

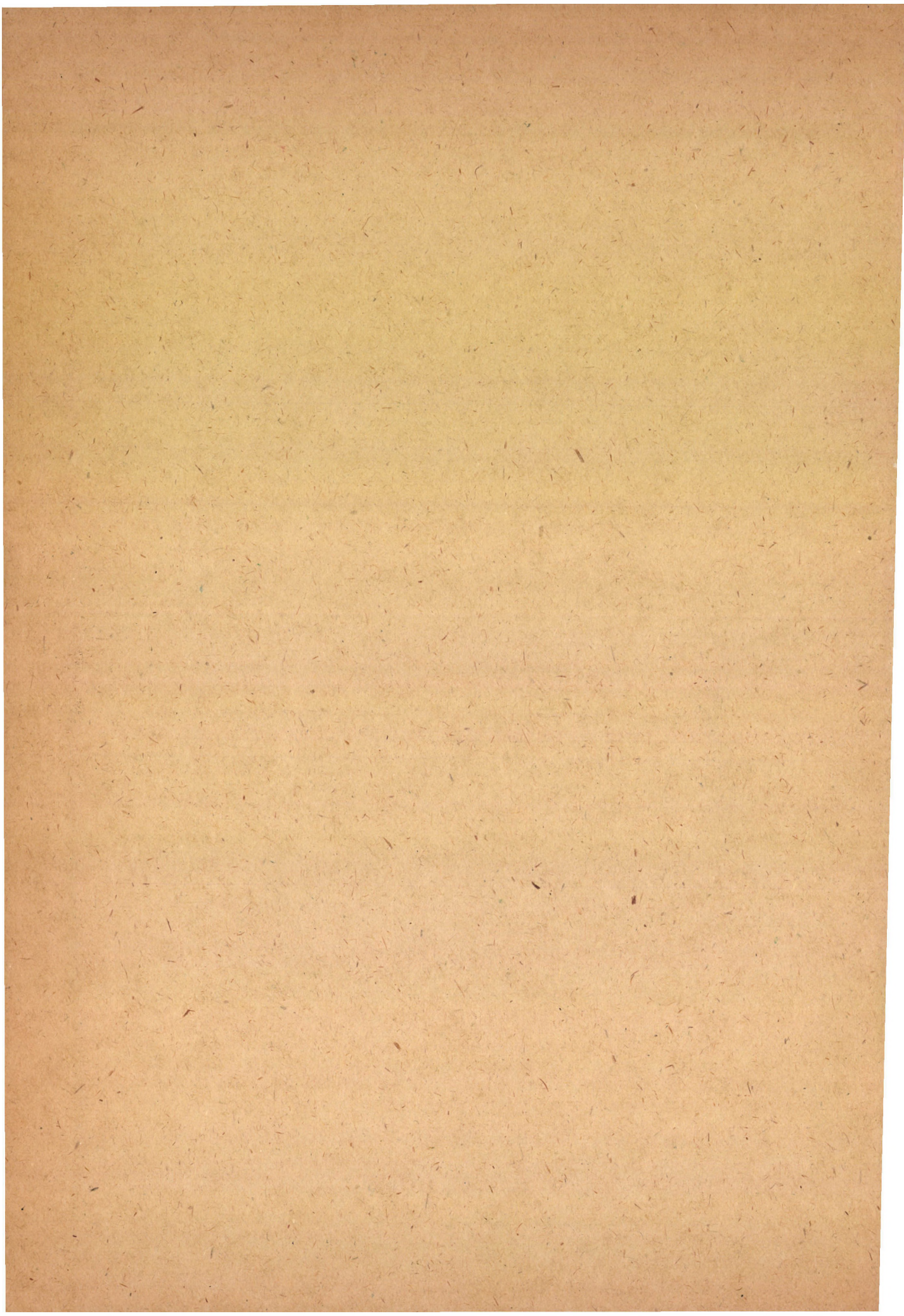
314.417

1964

COMPUTATIONAL LINGUISTICS

II

COMPUTING CENTRE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST, 1964



C O N T E N T S

EDITOR'S NOTE	3
S.ABRAHAM	
A Formal Study of Generative Grammars I.	5
L.DEZSŐ	
Некоторые вопросы языка-посредника (Some Questions of the Intermediate Language)	21
P.SGALL	
The Intermediate Language in Machine Translation and the Theory of Grammar	35
I.AGÓCS, D.VARGA	
Взаимосвязь элементов отношения русских и венгерских имён (The Interconnection between Relational Elements of Russian and Hungarian Nouns).	63
J.KLAUSZER	
Some Problems of Synthesis of the Hungarian Noun Forms	111
Gy.HELL	
The Two Levels of the Syntactic Analysis in the Mechanical Analysis of Russian Texts.	127
Gy.SIPŐCZY	
Some Semantic Aspect of the Machine Translation from Russian into Hungarian	159
I.BOTOS	
Опыт автоматической сегментации текста при МП с русского языка на венгерский (Some Experiments in Automatic Text Segmentation in Machine Translation from Russian into Hungarian).	179
APPENDIX	
F.KIEFER	
Mathematical Linguistics and MT in the Computing Centre of the Hungarian Academy of Sciences	203
F.PAPP	
Mathematical and Applied Linguistics at the University of Debrecen	213
M.STEIN	
MT Work by the Computer Elliott 803	219
BIBLIOGRAPHY of the Material Received up to January 1964.	223
CONTENTS	229

C O M P U T A T I O N A L L I N G U I S T I C S

II.

COMPUTING CENTRE OF THE HUNGARIAN ACADEMY OF SCIENCES

BUDAPEST, 1963

EDITOR

FERENC KIEFER

EDITORIAL BOARD

FERENC PAPP, JÁNOS S.PETŐFI,
GYÖRGY SZÉPE, DÉNES VARGA

Publisher: The Computing Centre of the Hungarian
Academy of Sciences

Address : Budapest, V., Nádor utca 7.

Felelős kiadó: Frey Tamás

Alak: A/4 Ivszám: 32,0

Megjelent 1964.március hóban, 500 példányban

64.2711 - Fővárosi Ny.V.16.t.(lehuzás)

EDITOR'S NOTE

The controversies called forth by the papers of Y.Bar-Hillel [1] and recently by A.G.Oettinger [2] about the future of MT are well known to all MT-workers. Unfortunately, these papers have not raised a public discussion in any journal or periodical as yet. All research workers in the field of MT must be interested in the discussion of the points of view expressed in the mentioned papers, because they are doubly vital for the further development of MT. Thus, all readers (MT-workers, linguists, mathematicians, logicians etc.) are requested to express their opinion about the prospects of MT in general, and from the economic point of view in particular, about problems that can be solved within the scope of economy and the most important obstacles (if any) to MT.

Beginning with the next issue of CL we start publishing the answers to these questions. Contributions should be sent to the editor.

- [1] BAR-HILLEL, Y., The Present Status of Automatic Translation of Languages, in Advances in Computers I. (ed. Franz Alt), Academic Press, New York 1960, pp.91-163

BAR-HILLEL, Y., Four Lectures on Algebraic Linguistics and Machine Translation, Hebrew University, Jerusalem 1963

- [2] OETTINGER, A.G., The State of the Art of Automatic Language Translation: An Appraisal, in Sprachkunde und Informationsverarbeitung 2, Oldenbourg Verlag, München 1963, pp.17-29

A FORMAL STUDY OF GENERATIVE GRAMMARS (I)

by S. Abraham

I.

1. The base of the study of generative grammars is the following system.

There are given a finite set of symbols (letters) V_p , called dictionary, which contains the identical element I (for which we have $IX=XI=X$ for every $X \in V_p$) and the boundary element $\#$, a finite set Σ of sequences in V_p , called initial words (all initial words begin and end with the boundary element), and a finite set F of rules of the form $\varphi \rightarrow \psi$, where φ and ψ are words (sequences in V_p). To apply a rule $\varphi \rightarrow \psi$ to the word $A_1 \omega A_2$ means to generate the word $A_1 \psi A_2$.

Imposing some additional conditions on the rules from F , we get different generative grammars.

This generative system (with Chomsky: generative grammar of type 0) works in the following way. We begin with a certain initial word, say $\# S \#$, and by applying the rules from F , we generate new words, till we generate a word $\# \omega_n \#$, to which no rules from F can be applied. The word ω' follows from the word ω if $\omega = \dots \varphi \dots$, $\omega' = \dots \psi \dots$ and $\varphi \rightarrow \psi$ is a rule from F . The sequence of words $\# S \#, \# \omega_1 \#, \dots, \# \omega_n \#$ is called a derivation in our system if each $\# \omega_i \#$ follows from $\# \omega_{i-1} \#$ ($i=1, \dots; \# S \# = \# \omega_1 \#$). The sequence of rules from F : f_1, \dots, f_n which are applied to obtain the given derivation is called the scheme of the derivation (and we shall write: $(f_1, \dots, f_n) (\# S \#) = \# \omega_n \#$).

If to $\# \phi_n \#$ no rule from F can be applied, $\# \phi_n \#$ is called a terminal word (and we have a terminal derivation).

The set of terminal words L_S is called the language generated by the given system.

Usually, the terminal words are sequences in a subset $V_{p,t}$ of the set V_p (called the terminal dictionary). The rules from F which at the right side have words in $V_{p,t}$ are called terminal rules.

2. On the basis of some linguistic considerations (generative systems are usually constructed for the study of natural languages) it is required that if $\rho = \varphi_1 A \varphi_2$ and $\psi = \varphi_1 \omega \varphi_2$, A should be a single symbol from V_p (and neither A or ω can be \dot{I})^x. This condition is an essential one, because if it holds, the system is recursively enumerable.

The rules $\varphi_1 A \varphi_2 \rightarrow \varphi_1 \omega \varphi_2$ can be of two types:

- (a) recursive rules, if $A \in \omega$, and
- (b) linear rules, if $A \notin \omega$

The rules of type (a) are right recursive if $\omega = \alpha A \beta$ and $\beta = \dot{I}$, $\alpha \neq \dot{I}$; left recursive if $\alpha = \dot{I}$, $\beta \neq \dot{I}$ and self-embedding if $\alpha \neq \dot{I}$, $\beta \neq \dot{I}$.

For the generative power of a system S it is highly important whether it contains or not recursive rules.

3. Between two systems S_1 and S_2 we can have the following relations:

^x If $\rho_1 = \varphi_2 = \dot{I}$ we have (Chomsky) a type 2 grammar; on the contrary a type 1 grammar.

$$(1) \quad V_{p1} \subseteq V_{p2}$$

$$(2) \quad \Sigma_1 \subseteq \Sigma_2$$

$$(3) \quad F_1 \subseteq F_2$$

and hence,

$$(4) \quad L_{S1} \subseteq L_{S2}$$

In such a case we say that S_2 includes S_1 , or that S_1 is enlarged to S_2 (and we write $S_1 \subseteq S_2$).

If (1) - (3) holds and $L_{S1} = L_{S2}$, then we say that the inclusion (enlargement) is a formal one. We have a formal enlargement, for instance, if $\Sigma_1 = \Sigma_2$ and the rules $F_2 \setminus F_1$ have no elements from F_1 at the left side.

4. The system constructed above is a semi-Thue system [1] or a semi-associative system [2].

Thus, knowing the interconnection between the associative (and semi-associative) systems and the normal algorithms [2], we can study the generative grammars in the terms of the theory of algorithms.

For this we shall first study in some details how a generative system works.

5. For the type 2 grammars, where the application of a rule is in no way restricted, we can agree to apply each rule successively as many times as it is possible.

If in such a grammar, during the terminal derivation

$$(1) \quad d = \omega_1, \dots, \omega_n \quad (\omega_1 \in \Sigma)$$

to each σ_i only one rule f_i can be applied, the terminal word σ_n can be obtained also by applying to σ_1 the/normal/algorithm

$$\alpha \left\{ \begin{array}{l} f_1 \\ \vdots \\ f_i \\ \vdots \\ f_{n-1} \end{array} \right. \quad (f_i (\sigma_i) = \sigma_{i+1})$$

If in (1) σ_i contains m symbols, and between f_1, \dots, f_n , we find a rule, f_k , at least $m+1$ times, then f_k is a terminal rule.

On the other hand, if at some step of the derivation (1) to the word σ_i we can apply the rules f_i and f'_i (no one of which is recursive), we have two derivations: (1) and (1') (with the scheme f_1, \dots, f'_i, \dots). If (1') is also a terminal derivation, then we obtain the terminal words σ_n and σ'_n . σ'_n can also be obtained by applying to σ_1 the algorithm

$$\alpha \left\{ \begin{array}{l} f_1 \\ \vdots \\ f'_i \\ \vdots \\ f'_{n-1} \end{array} \right. \quad ((1') \text{ has the scheme } f_1, \dots, f'_i, \dots, f'_{n-1})$$

The case when to the words $\sigma_{i_1}, \dots, \sigma_{i_e}$ of the derivation (1) can be applied, respectively, the rules $f_{i_1}, f'_{i_1}, \dots, f_{i_1}^{(k)}, \dots, f_{i_e}, f'_{i_e}, \dots, f_{i_e}^{(p)}$ can be handled in a similar way.

Instead of the algorithm α the algorithm can be considered that we obtain writing every rule in α only once, or the algorithm which we obtain completing α with all the rules from F which α does not contain.

In the case of type 1 grammars already it can be not quite irrelevant that we do apply a rule successively more than once or do not.

Such is the situation, for example, in the system

$$(1) \quad ABC \longrightarrow ABD$$

$$(2) \quad DDA \longrightarrow DDB$$

where if (1) is applied successively to ABCDABC we obtain ABDDBBBD. If we apply (1) and after that (2), we obtain ABDDBBC.

In such a case, and when a grammar contains rules $\varphi_1^A \varphi_2 \longrightarrow \varphi_3^A \varphi_4$ with one of the relations:

$$I \quad \varphi_1 = \varphi_2 = \varphi_3 = i; \varphi_4 \neq i$$

$$II \quad \varphi_1 = \varphi_2 = \varphi_4 = i; \varphi_3 \neq i$$

$$III \quad \varphi_1 = \varphi_2 = i; \varphi_3 \neq i; \varphi_4 \neq i$$

$$IV \quad \varphi_2 = i; \varphi_1 = \varphi_3 \neq i; \varphi_4 \neq i$$

$$V \quad \varphi_1 = i; \varphi_2 = \varphi_4 \neq i; \varphi_3 \neq i$$

we must be able to realise algorithmically the application of a rule as many times as we want.

For this we introduce the notion of a subscheme of the derivation's scheme: each subsequence of the scheme of the derivation is called a subscheme of the derivation. So, each scheme can be considered as the concatenation of a finite number of subschemes.

For instance, the scheme $f_1, f_1, f_2, f_3, f_4, f_1, f_5, f_6, f_2$ can be considered as consisting from the subschemes $f_1, f_1; f_2, f_3, f_4; f_1; f_5, f_6; f_2$ (we shall use the notation: $(f_1, f_1, f_2, f_3, f_4, f_1, f_5, f_6, f_2)(\mathcal{G}_i) = (f_1, f_1)(f_2, f_3, f_4)(f_5, f_6)(f_2)(\mathcal{G}_i) = \mathcal{G}_j$.

In the terms of the theory of algorithms the composition of algorithms corresponds to the concatenation of subschemes.

If we want to retain the properties of the rules from F for the rules which form the algorithm - result of such a composition, the Markovian way of composition of algorithms must be replaced by another one.

But first let us recall the properties of the rules from F :

(1) each rule re-writes at once only one symbol

(2) if $f(\mathcal{G}) = \mathcal{G}'$ then \mathcal{G}' must have the length (number of symbols) not less than \mathcal{G} .

To reach our goal, we shall use the following

Theorem: If $\mathcal{A}_{v_p}, \mathcal{B}_{v_p}$ are two normal algorithms, then, for every $\mathcal{G} \in \Sigma$, we have

$$\mathcal{G}(\mathcal{G}) = (\mathcal{A} \circ \mathcal{B})(\mathcal{G}) = \mathcal{B}(\mathcal{A}(\mathcal{G}))$$

where \mathcal{G} is an algorithm with the scheme

$$\left\{ \begin{array}{l} \bar{\xi} \eta \rightarrow \bar{\xi} \bar{\eta} \\ \bar{\mathcal{B}} \\ \xi \bar{\eta} \# \rightarrow \cdot \xi \eta \# \\ \bar{\xi} \rightarrow \xi \\ \mathcal{A} \\ \xi \rightarrow \bar{\xi} \end{array} \right.$$

where $\xi, \eta \in V_p$; $\bar{\xi}, \bar{\eta}$ are symbols put in one-to-one correspondence to the symbols from V_p (and different from them and between them); \bar{B} is the list of the rules of the algorithm B with every ξ changed to $\bar{\xi}$.

The rules of the algorithm \mathcal{G} have the properties (1) and (2) if \mathcal{A} and B have these properties.

If we want now to obtain algorithmically that a rule f_i should be applied n times succesively (and f_i can be applied succesively m times, $m > n^x$), we shall consider this as a subscheme, and proceed in the following way.

We shall consider the case when $n = 2$ (which can easily be generalised). If f_i has the form $\varphi_i^A \varphi_j \rightarrow \varphi_i \omega \varphi_j$, we choose two sets $V_p^{(1)}, V_p^{(2)}$ (of the same power as V_p), so that $V_p \cap V_p^{(i)} = V_p^{(1)} \cap V_p^{(2)} = \emptyset$ and $\xi_j^{(i)} \neq \xi_k^{(i)}$ when $j \neq k$. Then we realize the one-to-one correspondences $V_p \leftrightarrow V_p^{(1)}, V_p \leftrightarrow V_p^{(2)}$ so that $\xi_i \leftrightarrow \xi_i^{(1)}, \xi_i \leftrightarrow \xi_i^{(2)}$ ($\xi_i \in V_p, \xi_i^{(1)} \in V_p^{(1)}, \xi_i^{(2)} \in V_p^{(2)}$). The wanted algorithm will be the composition of the algorithms:

$$\mathcal{A}_1 \begin{cases} \eta \xi^{(1)} \rightarrow \eta^{(1)} \xi^{(1)} \\ \xi^{(1)} \eta \rightarrow \xi^{(1)} \eta^{(1)} \\ \varphi_1^A \varphi_2 \rightarrow \varphi_1 \xi_1^{(1)} \dots \xi_n \varphi_2 \end{cases}$$

$$(\omega = \xi_1, \dots, \xi_n; \eta, \xi \in V_p)$$

$$\mathcal{A}_2 \begin{cases} \eta^{(1)} \xi^{(2)} \rightarrow \eta^{(2)} \xi^{(2)} \\ \xi^{(2)} \eta \rightarrow \xi^{(2)} \eta^{(2)} \\ \varphi_1^{(1)} \xi_1^{(1)} \dots \xi_n^{(1)} \varphi_2 \rightarrow \varphi_1^{(1)} \xi_1^{(2)} \dots \xi_n^{(1)} \varphi_2 \end{cases}$$

(where $\varphi_j^{(1)}$ is the result of the substitution in φ_j all ξ by $\xi^{(1)}$).

$$\mathcal{A}_3 \{ \xi^{(2)} \rightarrow \xi \}$$

^x If f_i has the form $\varphi_1^A \varphi_2 \rightarrow \varphi_3^A \varphi_4$ with one of the relations I - V, $m = \infty$.

6. As we have seen, such an algorithm corresponds to each scheme of terminal derivation that has as its scheme a permutation of the rules from F , or that is the composition of such algorithms with algorithms regulating the number of the successive applicability of the rules from F (if it is the case).

So, we can attach to a generative grammar $[V_p, \Sigma, F]$ a set M of algorithms that corresponds to the terminal derivations in this grammar. We shall now consider the system $[\Sigma, M]$ and call the language generated by this $(L[\Sigma, M])$ the set of the words obtained by applying the algorithms from M to the words from Σ (if $\mu \in M$ and $G \in \Sigma$ then $\mu(G) \in L[\Sigma, M]$).

Evidently $L[\Sigma, M] = L[V_p, \Sigma, F]$.

Each word $\mu(G)$ from $L[\Sigma, M]$ is obtained by applying the rules from F to a word from Σ and is terminal in $[V_p, \Sigma, F]$.

Each terminal word G_T from $L[V_p, \Sigma, F]$ is word in $L[\Sigma, M]$ because it is obtained by applying some rules from F to a word from Σ , and, being a terminal word, the scheme of its derivation can be completed with all the other rules from F . But proceeding so, we have an algorithm from M which applied to the same word from Σ , gives G_T .

Taking in consideration that in the case of grammars type 2 and grammars type 1 without rules with one of the properties I-V, the set M contains only a finite number of algorithms, and that Σ has a finite character, we have the important result that L_M contains also a finite number of words. This means that such grammars are inadequate to generate natural languages, which, it is convenient to consider, have an open character.

In the case of grammars with rules for which one of the I-V properties holds, the consideration of the (finite) reunion

of sets $V_p, V_p^{(1)}, V_p^{(2)}, \dots$ leads only to a formal enlargement of the system, because the rules concerning the symbols $\xi^{(i)}$, do not contain at their left, elements from Σ .

In the case of such grammars, the sets M and $L[\Sigma, M]$ are enumerable.

II.

The above-described grammars turned out not to be quite fit for the description of natural languages: type 1 grammars give for some sentences quite an unintuitive structural description; type 2 grammars do not take in consideration the context, which often is desired. And generally it is not known whether natural languages can or cannot be generated by such grammars (called: immediate constituent grammars, IC-grammars). Starting from this unsatisfactory situation Chomsky proposed another (generative) grammar: the transformational grammar. A transformational grammar consists of two levels: an IC level, on which it is not required to generate all the sentences of the language under consideration; and a second level (transformational) containing transformational rules or transforms, which, applied to terminal sequences of the first level, generates all the remaining sentences of the language. For the transformational rules do not hold the restrictions of the IC level: (a) they can re-write at once more than one symbol; (b) they can consider the IC structure of the sequence to which they apply; (c) they can generate a sentence by being applied to more than one sequence.

The study of transformational grammars is not advanced because they are not satisfactorily formalized.

To arrive at such a formalization, we should consider first the transformational grammars in which the transformational rules can be applied to one sequence only.

The application of some operation to a sentence can have as result: (a) changing of the order of the symbols; (b)

adding of some symbols; (c) deleting of some symbols; or combination of (a), (b), (c).

These results can be obtained algorithmically also.

If from λYZ we want to obtain $ZY\lambda$, we can apply to λYZ , the algorithm

$$U \left\{ \begin{array}{l} XYZ \rightarrow \alpha YZ \\ \alpha YZ \rightarrow \alpha \beta Z \\ \alpha \beta Z \rightarrow \alpha \beta \sigma \\ \alpha \beta \sigma \rightarrow Z \beta \sigma \\ Z \beta \sigma \rightarrow Z Y \sigma \\ Z Y \sigma \rightarrow Z Y \lambda \end{array} \right. \quad (\alpha, \beta, \sigma - \text{different symbols, not contained in } V_p)$$

This can be generalized also for the case when between the symbols which should be interchanged there are more than one symbol.

We can algorithmically re-write a single XYZ as $ZY\lambda$ even in the case of a sequence containing XYZ more than once, proceeding like on p.5.

It is easy to obtain λUY from XY : we must apply to XY the algorithm

$$B \{ \lambda Y \rightarrow \lambda UY$$

To delete the symbol U from λUY , we must apply to λUY the algorithm

$$C \{ \lambda UY \rightarrow \lambda \wedge Y^x$$

The condition that a transformational rule should be applied only to sequences of the given structure (in IC) is equivalent with the condition that an algorithm should be applied then (and only then), when an other algorithm generates

^x and add to the morphophonematical rules some additional rules.

a given sequence of symbols. Every terminal sequence of the first level can be obtained by applying an algorithm α to an initial word. If in the inverse of α , α^{-1} we add, as the first substitution rule, the following

$$s \rightarrow \underbrace{|| \dots |}_n$$

where s is the structure required by the transformational rule, and n is the number of the symbols in s ($| \notin V_p$), this new algorithm α' , will re-write as a natural number, those and only those terminal sequences which have the structure s .

The way we proceed is the following.

If M is the (finite) set of algorithms which generates the language L , we construct the set \bar{M} of algorithms, so that to the inverse of each algorithm from M , we add as first substitution rule, the following:

$$s \rightarrow \underbrace{|| \dots |}_n$$

We denote by $B_{n,s}$ the algorithm

$$\underbrace{|| \dots |}_n \rightarrow s$$

and by \tilde{I} the algorithm which produces the changes required by the considered transformational rule. Then a transformational rule can be replaced by the algorithm

$$G = \alpha' \circ B_{n,s} \circ \tilde{I}$$

If G has no the structure s , we have

$$G(G) = G$$

The algorithm \tilde{I} is, in fact, the composition of algorithms which produce changes of the type (a), (b), (c). So we can write: $\tilde{I} = \tilde{I}_a \circ \tilde{I}_b \circ \tilde{I}_c$. So that

$$(1) \mathcal{G} = \overline{\mathcal{A}}' \circ \mathcal{B}_n, s \circ \mathcal{I}_a \circ \mathcal{I}_b \circ \mathcal{I}_c$$

By introducing the symbol $|$, we made only a formal enlargement of the system, because $\varepsilon \notin \Sigma$ and $\text{لـنـبـل} \notin \Sigma$, for $e \geq 1$.

If we consider the already formulated (non-formally) transformational rules, we see that they formulate the changes in the terms of the order of symbols which form the words.

Such a formulation of the algorithm \mathcal{I} can be obtained if we consider in V_p a symbol $+$, so that between every two symbols (from V_p) we have $+$. Then \mathcal{I} must have as a first component the algorithm

$$\mathcal{R} \begin{cases} \xi + \rightarrow x_i + \\ x_n + \xi \rightarrow x_n + x_{n+1} \end{cases} \quad \begin{matrix} \xi \in V_p; x_i \notin V_p \\ n = 1, 2, \dots \end{matrix}$$

The other algorithms which form \mathcal{I} , are then formulated in terms of x_i , and we add $\overline{\mathcal{R}}$ (the inverse of \mathcal{R}).

Now we can define a transformational grammar, as a system $[\Sigma, M, \mathcal{G}]$.

The algorithms of \mathcal{G} have not the properties of the algorithms from M (which form the first level):

- (a) the algorithms from \mathcal{G} have not the properties from p.6.
- (b) the algorithms from M apply to words from Σ while the algorithms from \mathcal{G} to the terminal sequences of the first level (we shall denote their set by L_{IC}).

The second difference can be removed in the following way.

If $t \in L$ and $t = \mathcal{G}_i(\sigma_j)$ where $\sigma_j = M_j(\sigma_1)$ with $\sigma_1 \in \Sigma$, we have

$$t = \mathcal{G}_i(M_j(\sigma_1)) \quad \text{or}$$

$$t = (M_j \circ \mathcal{G}_i)(\sigma_1).$$

Now, if we denote

$$M_j \circ \mathcal{G}_i = \mathcal{G}'_{j,i}, \text{ we have}$$

$$t = \mathcal{G}'_{j,i}(\mathcal{G}_1) \text{ and } L = M(\Sigma) \cup \mathcal{G}'(\Sigma)$$

(a) takes place because $\overline{\alpha'}$ and $\overline{\kappa}$ has not the mentioned properties (the inverse of an algorithm has not these properties).

If we could construct the inverse of an algorithm with the mentioned properties, so that these properties would remain, the transformational grammars we discussed would be IC grammars.

But, even so, it can be shown that for each system $[\Sigma, M\mathcal{G}']$ it is possible to construct a system $[\Sigma, M]$ so that $L[\Sigma, M\mathcal{G}'] = L[\Sigma, M]$.

Indeed, every word $t \in L \setminus L_{ic}$, obtained by applying a transformational rule (RT) to a terminal sequence σ , has a given s_t structure (in IC). So we have a derivation where s_t is the first word and t the terminal one, and s_t is not terminal in any derivation which begins with a word from Σ . But s_t can be obtained from s (the structure mentioned in the formulation of RT by the operations (a), (b), (c)).

But now, s is obtained by an algorithm with the required properties; (a), (b), (c) can also be obtained with algorithms with the same properties, and the composition of algorithms maintain these properties, so we can obtain t by applying to a word from Σ an algorithm with the required properties. This means, that the IC level of the discussed transformational grammar, can be completed to an IC grammar which generates also t . Such a completion consists in completing the rules of F with the substitution rules needed to obtain s from s_t .

If we proceed in the same way with all the transformational rules, then we have the result that for every transformational grammar $[\Sigma, M, G']$, that re-writes only one sequence, we can construct an IC grammar $[\Sigma, M]$ so that $L[\Sigma, M, G'] = L[\Sigma, M]$.

BIBLIOGRAPHY

- [1] DAVIS, M., Computability and unsolvability, NY, Mc Graw-Hill, 1958
- [2] MARKOV, A. A., Teoria algorifmov, Trudi mat. inst. AN SSR, XLII, Moscow, 1954
- [3] CHOMSKY, N., On certain formal properties of grammars, Information and Control 2, pp. 137-167, 1959
- [4] CHOMSKY, N., Three models for the description of language. IRE Trans. on Inform. Theory IT-2, pp. 113-124, 1956

Л. Дәже

Машинный перевод может осуществляться путем перевода с одного языка на другой, но возможно и то, что текст входного языка переводится на язык-посредник /сокращенно ЯП/ и с ЯП на выходной язык. При том для посредничества могут служить не только лингвистические средства, но и иные, например, формальный язык математической логики. В данной статье мы ограничимся рассмотрением первых. В качестве ЯП могут выступать естественные языки, искусственные языки или специальный ЯП, созданный для этой цели. ЯП нередко называется интерлингвой, что нельзя считать удачным, так как этот термин уже давно обозначает определенный искусственный язык. В дальнейшем главное внимание обращается на язык, специально созданный для цели языка-посредника МП, но коснемся и возможности употребления естественных и искусственных языков в качестве ЯП.

1.

Первый элементарный вопрос, который возникает при рассмотрении ЯП: стоит ли употреблять ЯП. Этот вопрос вовсе не второстепенный, ведь создание ЯП требует немало работы и материальных средств, которые можно было бы обратить и на разработку бинарных алгоритмов, приносящих более верные проценты.

Несомненно, что при применении ЯП требуется намного меньше алгоритмов.⁺ Однако данный аргумент с полным правом оспаривается противниками ЯП, так как он не считается возможностями конкретного создания алгоритмов МП, ведь чем больше алгоритмов составляется, тем легче составление новых алгоритмов. Правда,

⁺ - Ср. Н.Д.Андреев, Машинный перевод и проблема языка посредника. ВЯ 6:5 стр. 117-121.

этот аргумент можно вывернуть наизнанку: алгоритмы, присоединяющиеся к ЯП тоже составляются легче с увеличением их числа. Но данный вопрос пока что отложим в сторону. Алгоритмами ЯП подробнее займёмся ниже, а прежде нуждается в обсуждении другое, более важное обстоятельство: базис ЯП.

Когда мы переводим текст входного языка на ЯП, по существу, мы получаем "полуфабрикат": ряд информации входного языка, выраженных в категориях ЯП.

Перевод может осуществить лишь тот, кто имеет соответствующий алгоритм, основывающийся на знании входного языка. На ЯП могут переводить: /1/ переводчики входного языка, /2/ переводчики какой-то другой страны, если они осуществляют свой бинарный перевод через ЯП. Переводчики выходящих языков могут получить серию информации ЯП только у них, потому что они, без знания входного языка не в состоянии анализировать переводимый текст по категориям ЯП. Ясно, что применение ЯП возможно лишь в том случае, если между соответствующими сторонами имеет место интенсивный, организованный обмен научными и техническими информацией.

При работе по ЯП мы исходим из возможности обмена опытом между социалистическими странами на высоком уровне. Поэтому нам предстоит в качестве первой задачи создание такого ЯП, который может удовлетворить этим требованиям. Конечно, никто не может отказаться от плана ЯП, основанного на официальных языках всех стран, если имеются реальные условия для этого. Система ЯП, базирующегося на меньшем количестве языков, должна быть такой, чтобы с увеличением числа языков, составляющих его основу, можно было его дополнять.

Итак, мы дошли до вопроса о том, на каких языках должен основываться ЯП, служащий для обмена опытом социалистических стран.

1./ Языки социалистических стран.

- 2/ Языки несоциалистических стран с развитой промышленностью /германская и романская группы языков, японский/.
- 3/ Важные азиатские и африканские языки /индонезийский, хиндустани, арабский, суахели/.

При конкретном осуществлении в зависимости от возможностей надо было бы начать с более узкой основы. С лингвистической точки зрения на определенной ступени работы было бы целесообразнее разделить основные языки на языки входно-выходные и только входные. Так, например, языки второй группы были бы только входными при ЯП, употребляемом между социалистическими странами. Мы должны подчеркнуть, что создание ЯП на более узкой основе может быть целью только на начальной стадии развития работы.

Далее. ЯП должен содержать те категории, которые сделают возможным полный перенос информации входного языка на выходной язык. Категории ЯП, несмотря на то, что они построены на формальном анализе конкретных явлений, должны быть абстрактными. Надо найти то общее, что скрывается за конкретными явлениями отдельных языков. Такие категории ЯП снова будут превращены в конкретные формы выходных языков.

Для успешного перевода при помощи ЯП мы должны располагать многоязычными микрословарями, которые включают в себя общую научную и техническую лексику и словари отдельных отраслей науки и техники. В лексике науки и техники общей основой слов является соответствующее понятие, которое выражается по-разному в различных языках. В научно-технических терминах мы будем встречать меньше трудностей, но мы сталкиваемся с большими проблемами при соотношении общей лексики.

Машина рассматривает формальные признаки данного языка, при помощи которых можно установить так называемые "грамматические значения" слова, как: число, лицо, время, наклонение и др. Данные ЯП должны содержать информации об этих категориях для того, чтобы можно было найти соответствующие категории и

формы их выражения в выходных языках. Та же задача стоит перед нами и при бинарном переводе, только при переводе с помощью ЯП одному языку соответствуют не один, а много языков, что приводит к значительному увеличению трудностей. Разумеется, трудности не увеличиваются соответственно числу языков, ведь большинство категорий отдельных языков является общим.

При бинарном переводе главной трудностью является приобретение информации, необходимых для того, чтобы выражать те особенные категории выходного языка, которые неизвестны входному языку. Так, например, при переводе с русского на венгерский: венгерские объектное и субъектное спряжения, разделяемые префиксы глаголов и др. Ясно, что если перевод осуществляется не на один, а десять языков, число таких проблем увеличивается. Это и является ахилесовой пятой перевода с ЯП.

К формированию особенных категорий выходных языков ведут два пути. Один из них заключается в том, что при анализе входного языка стараются установить все информации, нужные для всех выходных языков. Значит, ЯП будет содержать все категории, которыми располагают выходные языки. Это приводит к значительному усложнению системы ЯП и удлинению анализа. Идя по другому пути, стараются исключить из системы ЯП все то, что переводчики выходного языка самостоятельно могут установить на основе полученных информации. Какими именно будут эти информации, можно установить лишь при конкретном использовании ЯП. Несомненно, что состав категорий нельзя установить заранее даже при таком сопоставлении особенностей отдельных языков, которое предложено Н.Д.Андреевым⁺. Если мы желаем получить хорошо составленные предложения, то ряд информации ЯП должен содержать все необходимое для этого. Если в венгерском языке для правильного составления предложений мы нуждаемся в употреблении артикля, то ЯП должен располагать информацией, необходимыми для определения артикля даже тогда, если языки, имеющие эту категорию, составляют меньшинство. Категория артикля

⁺ - Ср. Машинный перевод и проблема языка посредника.
ВЯ 6:6 стр. 121.

может отсутствовать в ЯП лишь тогда, если на основе полученных информации переводчики выходного языка могут установить, когда какой артикль надо употреблять.

Здесь представляются две возможности избежания этих трудностей, но ни одна из них не решает проблемы удовлетворительно. Первая: при составлении предложений выходных языков мы оставляем в стороне все то, что причиняет трудности при переводе с ЯП, но не мешает пониманию текста. Так, предложения будут хорошо понятны, но грамматически неправильно составленные. О другой возможности уже упоминалось выше: основные языки разделяются на входно-выходные и только входные языки. Категории только входных языков не учитываются при анализе данного входного языка, тем самым упрощается анализ. Такие упрощения можно считать лишь переходными.

Вернемся, однако, к главному вопросу. Рассмотрим на конкретном примере вопрос об особенных категориях выходных языков, не входящих в ЯП.

Субъектное и объектное спряжения являются особенностями венгерского языка. Необходимо ли, чтобы ряд информации, полученный через ЯП, содержал специальные данные о том, какое спряжение употребляется при переводимом глаголе? На наш взгляд нет необходимости в том, чтобы эта особенность венгерского языка, неизвестная другим основным языкам, получила бы место среди категорий ЯП. Объектное спряжение употребляется в следующих четко определенных случаях: если объект (а) имя собственное, (б) имя с определенным артиклем, (в) личное местоимение 3-го лица, (г) указательное местоимение, замещающее имя существительное, (д) возвратное местоимение, (е) взаимное местоимение egymást, (ж) обобщительное местоимение mind, valamennyi, (з) местоимение или порядковое числительное с -ik в конце слова, (и) конструкция с определением, выраженным местоимениями, перечисленными в пунктах (г) (ж) (з); (к) имя с личным притяжательным окончанием -é или с притяжательной прилепой, (л) инфинитив, (м) придаточное предложение (ср. *Mai magyar nyelvrendszere. Leiró nyelvrendszere. Budapest, 1962. стр. 158-160.*) Вот все, за исключением од-

ной, элементарные информации, которые обязательно содержатся в тексте. Впрочем, этот вопрос возникает и при бинарном переводе с русского на венгерский, так что опытом, приобретенным при его решении, мы можем пользоваться и в работе с ЯП.

Здесь возникает вопрос, который мы должны рассмотреть: каково соотношение работ по бинарному переводу и по ЯП?

По нашему мнению, перевод посредством ЯП является лишь одним из вариантов МП, который не делает излишней разработку бинарных алгоритмов. К этому можно добавить, что в первую очередь мы считаем необходимым разработать важнейшие бинарные алгоритмы. В ходе работ над бинарными алгоритмами будет выяснен ряд вопросов, без решения которых невозможно применение ЯП. Так, мы увидим, наличие каких информации необходимо для выражения особенностей венгерского языка при соединении венгерского языка к ЯП. Эти вопросы являются проблемами, подлежащими решению и при МП с русского на венгерский. Вместе с тем это означает, что реальные условия осуществления ЯП создаются в ходе разработки бинарных алгоритмов. Теоретически можно представить схему ЯП, но ее реализация зависит от того, можем ли найти машинные процедуры, нужные для выражения релевантных особенностей отдельных языков.

Подытожим вкратце сказанное о ЯП машинного перевода. Для применения ЯП требуется хорошо организованное тесное сотрудничество соответствующих стран, поэтому в начале мы можем ставить себе целью создать ЯП, основанный на ограниченном круге языков, вместе с тем такое задание является более реальным и легче разрешимым. Данные, сообщенные через ЯП, должны содержать все информации, достаточные для образования правильно составленных предложений на выходном языке. С другой стороны, ЯП должен содержать лишь те информации, которые необходимы для осуществления этого. Так, при помощи формального анализа основных языков, проведенного на высоком уровне, мы должны достичь, чтобы ЯП был свободен от всех категорий, выражение которых возможно на основе элементарных информации, получаемых

из всех языков. В этой области исследование ЯП может рассчитывать на серьезную помощь со стороны работ по бинарному МП.

2.

Среди категорий ЯП мы находим (1) такие элементарные, имеющиеся во всех языках, основные категории, как число, лицо, время и др., и (2) минимальное число категорий, характерных лишь для отдельных языков, например, артикль. Вопрос в том, как получить эти категории? Представляются два пути. Первый: исходить из регистрации категорий языков и, сопоставляя их, образовать систему категорий ЯП. Другой: исходя из сравнительно немногочисленных элементарных категорий, имеющих во всех языках, расширять их до тех пор, пока не будет образована оптимальная система категорий. Каким бы путем мы ни пошли, до решения, результат может быть только один. При этом мы должны учесть все особенности основных языков, независимо, от того, исходим ли мы из более узкого или широкого круга явлений.

Имея в виду выше сказанное, мы хотели бы вкратце познакомить читателя с грамматикой ЯП чехословацкой группы по МП. Это одна из первых конкретных попыток создания лингвистическими средствами специального ЯП. Грамматика ЯП пражской группы /"Entwurf der Grammatik einer Interlingua(IL)"/ опубликована как рукопись /сокращ.ИЛ/. До появления более подробного сообщения о чехословацком проекте ЯП, на наш взгляд, было бы преждевременно подвергать его детальной критике, поэтому мы только представим главные черты проекта.⁺ Пункты чехословацкого предложения служат для анализа текста входного языка, при помощи их из системы грамматических значений, общих для многих языков, мы должны установить те, которыми мы воспользуемся при создании форм выходного языка. Первые три пункта /Hilfangaben/ служат для локализации отдельных слов. Незнаменательные слова не

+ - После сдачи рукописи мы получили статью П.Сгалла: Prevodni jazyk a teorie gramatiky. Slovo a slovesnost. XXIV. 114-128, так что мы её не могли учесть.

получают отдельного номера, а рассматриваются как части знаменательных слов. Пункт 4 (Semoglyph) сообщает кодовый номер значения слова. Пункт 5 (Tektoglyph) показывает, от которого слова синтаксически зависит данное слово. Пункт 6 (Negation) требует данных о негации в предложении. Пункт 7 (Morphologische Semantik) служит для установления морфологических значений отдельных слов: (a) времени (b) релятивного времени, (c) вида глагола, (d) наклонения, (e) числа (каждый пункт подразделяется). Подробно разработан подпункт (f), в котором рассматриваются морфологические значения имен и придаточных предложений /напр., средство, цель, часть целого и др./, здесь находится и система значений предлогов и ряд данных, нужных для определения имен. Подпункт (g) (Bestimmtheit bei Substantiven) способствует установлению артикля (h) (Hervorhebung) занимается функциями частиц, подпункт (i) (Steigerung) рассматривает образование степеней сравнения. Пункт 8 (Koordination) служит для анализа сочинительной связи. Пункт 9 (Semantisches Satzglied) устанавливает агент, пациент предложения и др. независимо от того, какова его синтаксическая функция, что дает возможность, например, превратить страдательную конструкцию в действительную. Последний пункт 10 (Semantische Wortart) требует установления прилагательных, образованных от существительных; существительных, наречий, образованных от прилагательных; отглагольных существительных и др., так как в ходе перевода они могут быть изменены в другую часть речи, например, вместо соответствующего отглагольного существительного может быть употреблен глагол.

К проекту приложен ИЛ анализ короткого английского текста. Чтобы дать и конкретное представление о переводе посредством ИЛ, рассмотрим перевод первого предложения с английского на ИЛ и затем на венгерский.

Английское предложение: Electrical (1) conduction (2) occurs (3) in so-called (4) insulators (5) at high (6) temperatures (7). Цифры, стоящие в скобках, обозначают место слова, занимаемое в предложении. Незнаменательные слова не имеют особых цифр. Следующая таблица показывает информации, полученные в ходе анализа.

1	2	4	5	6	7							8	9	10
						a	b	c	d	e	f	g	h	i
1	1	107	2										D	Adj.
	2	213	3								Begr.		Ag.	V
	3	252	-			Präs. g.		dur.	Aus. sage				Pr.	V
	4	266	5										D	V
	5	46	3							Pl.	Ort. 3b	Unb.	D	S
	6	112	7										D	Adj.
	7	76	3							Pl.	rease Bed.	Unb.	D	S

Данные столбца 1 служат для локализации предложения, данные в столбце 2 для локализации слов внутри предложения. Столбец 4 дает кодовые номера слов в машинном словаре. В 5 находим указание на управляющее слово. Функции остальных столбцов и подразделений столбца 7 см. выше. Мы ограничиваемся лишь объяснением сокращений. Сокращения столбца 7 следующие: Präs(ens) g(enerelles) = общее настоящее время; dur(ativ) = дуративный глагол, Aussage = изъяснение, для характеристики глагола. Для характеристики существительного: Pl.= мн.ч., Ort 3b= обозначение места 'в чем-н?', real. Bed(ingung) = реальное условие, напр. венг. esetén; Begriffsbestimmtheit = определенный или неопределенный /артикл/, употребление которого зависит от самого слова; Unb(estimmtheit) = неопределенный /артикл/. Сокращения столбца 9 (Semantisches Satzglied): D = детерминант, Ag.= агент, Pr. = предикат, а столбца 10 (Semantische Wortart): Adj.= от прилагательного, V = отглагольный, S = от существительного. В ходе перевода на венгерский язык делается следующее: берем соответствующие венгерские слова из словаря, и на основе данных информации проводим необходимые в венгерском, или по данным информации возможные операции для адаптации. Порядок слов остается неизменным, так как относительно этого нет особых указаний. Мы не анализируем отдельно каждое слово, а только записываем слова, взятые из словаря, и затем объясняем те случаи, в которых наблюдается существенное изменение.

Венгерский текст: Elektromos (1) vezetés (2) létrejön (3) ugynevezett (4) insulatorokban (5) magas (6) hőmérséklet esetén (7). Нуждается в объяснении лишь окончание -ban слова (5), что точно соответствует значению, которое обозначается Ort. 3b в грамматике ИЛ, для выражения reale Bedingung в венгерском служит esetén. Образование остальных форм легко можно понять по данным таблицы.

Грамматика ИЛ содержит указание на употребление артикля, но не служит точкой опоры для определения порядка слов, не занимается вопросом субъектных и объектных спряжений и др. Итак, если мы хотим получить правильно составленные венгерские предложения, информации, необходимые для этого, должны содержаться в данных ИЛ, или на основе имеющихся информации, мы должны определить их сами. В отдельных случаях это может удаться, например, при определении субъектного или объектного спряжений.

Чехословацкий проект близок к тем соображениям, которые изложены в работах Н.Д.Андреева "Машинный перевод и проблема языка-посредника"⁺ и "Linguistic Aspects of Translation". Нашей задачей не является подробное сопоставление этих работ, мы хотели бы только более подробно остановиться на некоторых вопросах.¹⁾ Пражский проект не занимается установлением порядка слов, чему Н.Д.Андреев уделяет внимание.⁺⁺²⁾ Не совсем ясно, каково отношение пражской группы к тому пониманию принципа конгруэнтности, которое изложено Н.Д.Андреевым⁺⁺, и по проекту нельзя видеть, приняла ли его пражская группа. Из статьи Н.Д. Андреева видно стремление автора создать такой искусственный ЯП, который можно четко охарактеризовать, а по проекту ИЛ нельзя установить, является ИЛ закрытой системой или нет.

3.

Проект пражской группы был составлен на основе сравнительного анализа многих языков. Однако можно, как упомянуто выше,

+ ВЯ 1957:5.

++ Ук, соч. 121.

исходить из уже имеющегося языка. Рассмотрим данный вопрос на примере искусственного языка интерлингва.

Вопросом о возможностях МП с интерлингвы на основные языки, главным образом на английский, занимается статья Дж. Дарлингтона "Интерлингва и МП, Дискуссия".⁺ Автор рассматривает, действительно возможен ли механический перевод текста, написанного на интерлингве, на основные языки, как это утверждает А. Годэ в своей статье "Система знаков в интерлингве".⁺⁺ Анализируя конкретные примеры, Дж. Дарлингтон установил, что выражения интерлингвы часто непереводимы дословно на английский, в виду того, что данная конструкция имеет другое значение в английском, или она просто непонятна.⁺⁺⁺ Другой трудностью при переводе с интерлингвы является многозначность грамматических слов интерлингвы. Во избежание этого автор старается с помощью дополнительных элементов сделать однозначным для основных языков предлоги, союзы и местоимения интерлингвы.⁺⁺⁺⁺ Наконец, автор представляет читателю исправленные тексты анализированных примеров и их английский перевод.⁺⁺⁺⁺⁺ Статья Дж. Дарлингтона, хотя и не раскрывает всю сложность вопроса, но убедительно доказывает, что интерлингва не способна без существенных изменений быть посредницей даже между основными языками.

Однако нас интересует интерлингва как возможный язык-посредник МП. Для этого проводим следующие опыты: текст, написанный на интерлингве, переводим на русский и на венгерский, потом предположим, что его надо было бы перевести с русского на венгерский через интерлингву, и посмотрим, как справится интерлингва с этим заданием. Для большей вероятности текст напишем сначала на русском языке, потом на интерлингве и проанализи-

⁺ - J. Darlington, Interlingua and MT, a Discussion. Mechanical Translation. Vol. 7. No 1, 2-7.

⁺⁺ - A. Gode, Signal System in Interlingua. Mechanical Translation, Vol. 2 No 3.

⁺⁺⁺ - Ук. соч. 2-3.

⁺⁺⁺⁺ - Ук. соч. 4-5.

⁺⁺⁺⁺⁺ - Ук. соч. 6-7.

руем возможность его перевода на венгерский язык. Текст взят из руководства по интерлингве.⁺

Русский текст: Если мы примем такой словарь и сократим грамматику до минимума, как в китайском, мы имеем максимально международный и легкий язык, которым является латынь, употребляемая в языке науки, без флексии и без бесполезного препятствия и трудности грамматики. (Мы стремились к тому, чтобы русский текст был бы близок к оригиналу.)

Текст на интерлингве: Si nos adopta talé vocabulario et reduc grammatica ad minimo, sicut in sinense, nos habe maximo internationale et facile lingua, que es latino vivente in lingua de scientia, sine flexione et sine inutile impedimento et difficultate de grammatica.

Оставим в стороне проблемы перевода с русского языка на интерлингву, ибо следует лишь упростить более сложную систему, Рассмотрим вопросы перевода на венгерский язык.

Венгерский перевод с интерлингвы: Ha mi átves(s)zük(ünk) (az) ilyen(olyan) szókinccset és (le)redukáljuk(unk) (a) nyelvtant (a) minimálisra(nak) mint (a) kínaiban, mi bírunk(juk) maximális(an) nemzetközi és könnyű nyelvet, amely van (a) latin élő (a) tudomány(nak a) nyelv(é)ben flexió nélkül és (a) nyelvtan(nak a) szükségtelen akadály(a) és nehézség(e) nélkül.

Порядок слов интерлингвы сохранен, только порядок слов при-
тяжательной конструкции и послелога был изменен в соответствии с правилами венгерского языка.

Прежде всего мы должны признать, что возможен определенный элементарный перевод через интерлингву. Однако интерлингва как ЯП не содержит информации, например, об употреблении артикля, так что артикли поставлены нами в скобках. Не находим указания и относительно того, употребляется субъектное или объектное спряжение, не говоря уже о других специфических венгерских проблемах. Без артиклей нельзя решить и вопроса об употреблении субъектного и объектного спряжений. Правда, ар-

⁺ - Primo libro de interlingua. London, 1931. стр.1.

тиклей не было и во входном языке, но интерлингва не цодержала бы информации об артиклях даже тогда, если они были бы входном языке, например, при переводе с английского. Для того, чтобы интерлингва могла выполнить функцию ЯП машинного перевода, необходимо было бы изменить ее структуру в соответствии с особенностями основных языков.

Конечно, рассматриваемые примеры не могли показать всю проблематику вопроса, это и не могло быть нашей задачей, мы хотели дать представление только о некоторых, на наш взгляд, важных аспектах применения ЯП.

4.

Остается участь еще одну созможность: возможность применения естественного языка в качестве ЯП.

Нет необходимости доказывать, что можно выбрать один мировой язык, и статьи, представляющие интерес для многих стран, прежде всего переводить на этот язык. Разумеется, проблемы МП являются общими безотносительно к тому, переводится ли данный текст через искусственный ЯП, с естественного языка или при его посредстве. Избежать их нельзя. Кроме того, и в последнем случае мы должны справиться с трудным заданием в области составления алгсритмов. Отдельные страны должны располагать не только алгоритмами для перевода, например, с русского, но и на русский.

Здесь мы должны коротко коснуться и другого вопроса. ЯП машинного перевода существенно отличается от естественных, и даже от искусственных языков тем, что у него нет своей собственной формальной "материальной" стороны: он не имеет звуков, на место слов стоят номера кода понятий. Подлежит разъяснению, какой будет его минимальная лексическая семантика. Такой ЯП собственно не является языком, а только собранием особенностей основных языков, релевантных с точки зрения перевода. Он не

представляет собой закрытую систему, а может дополняться при включении новых основных языков и может быть изменен в соответствии с развитием наших лингвистических знаний.

- 0 -

В нашей статье мы старались показать две возможности создания ЯП. Из концепций о ЯП мы более подробно представили только проект пражской группы. О концепции Н.Д.Андреева говорилось лишь столько, сколько было необходимо в связи с ИЛ. Мы не упомянули о концепции И.А.Мельчука, которая была изложена в статье "Работы по машинному переводу".⁺ И.А.Мельчук отмечает два пути создания ЯП. Идя по первому пути, надо составить определенный искусственный язык, имеющий свою лексику, морфологию и синтаксис,⁺⁺ "Другой путь - рассматривать язык-посредник только как систему соответствий между реальными языками."⁺⁺⁺ В дальнейшем И.А.Мельчук излагает эту концепцию. Насколько можно судить по краткому изложению, второй путь определенно имеет общие черты с тем, что представлено в нашей статье. Мы имеем в виду в первую очередь то, что ЯП не должен быть закрытой системой, значит, каким-то искусственным языком, а системой релевантных категорий основных языков. Кроме того, статья И.А.Мельчука содержит ряд таких утверждений, которые неоспоримы, с какого направления бы ни подходили к проблематике составления ЯП. Другая весьма интересная статья И.А.Мельчука "К вопросу о "грамматическом" в языке-посреднике"⁺⁺⁺⁺ тоже рассматривает вопросы, затронутые и в нашей статье, только в другом плане, и ее анализом следовало бы заниматься особо.⁺⁺⁺⁺⁺

+ - Вестник Академии Наук СССР. 1959 г. стр. 42-47.

++ - Ук. Соч. 44-45.

+++ - Ук. Соч. 45.

++++ - Машинный перевод и прикладная лингвистика. 4:25-45.

+++++ - О вопросах ЯП говорилось в выступлениях Д.Калмара Ф.Паппа, Д.Сепе на конференции "Вопросы математической лингвистики" в Будапеште в марте 1962 года, материал которой находится в печати.

THE INTERMEDIATE LANGUAGE IN MACHINE TRANSLATION
AND THE THEORY OF GRAMMAR [1]

by

P. Sgall (Prague)

1. The possibility of a formal description of natural languages has been considerably increased by the achievements of the algebraical theory of grammar; [2] it is desirable to use these achievements directly also in machine translation.

2. It is necessary then to pay attention to the complicated relation between grammar and semantics. MT has to work with semantics, but an adequate theory of the semantics of natural languages cannot be constructed without large-scale applications of empirical approaches. [3] For such approaches we could find a number of contributive suggestions in the works of older linguists, [4] which should be elaborated systematically and tested by practical application on large texts, as it will be necessary in the preparation of the MT.

3. The grammar of an intermediate language (IL) in MT [5] should be constructed in such a way as to prevent the algorithms of natural language-to-IL translation from being more complicated than the algorithms for binary translation. Two points seem to be important in this respect: a) the grammar of IL should have a simple form, b) there should be no homonymy or synonymy (or as little of both as possible).

a) The first requirement can be fulfilled if we have in view a grammar which "bounds the language on the outside", [6] i.e. which does not exclude all "non-grammatical" strings (we can assume that there are none in the text to be translated, or that they were excluded by the algorithm of analysis of the input language).

b) As to the second point, the relation of "form and function" (or the semantic relation) in the grammar has to be con-

sidered. [7] The "tectogrammatical" and "phenogrammatical" elements have to be distinguished and only the former should enter the grammar of IL. Such classification of the elements of natural languages appears to be a crucial question. The relation of "form and function" seems to have several degrees. Following Mathesius, Havránek and other members of the Prague school, we can see the "highest" level of syntax in such relations as that of an action to its actor (agens) and to its "object" (or a necessary complement), the relation of an adjunct (or free complement) to its "nucleus", and further in the relations of coordination and apposition. Such units as subject, predicate, attribute etc. are then only means of individual languages "to represent" the functions of the higher level; e.g. the agens can have the form of a subject (in an active construction) or of an adverbial (in the passive construction) in the European languages, but other forms in the ergative languages etc. - In morphology it is possible to work with such units as Actual Present, Gnostic (General Tense), Past, ... Cause, Aim, Location (a. from, b. into, c. in...).

We naturally do not mean to say that such syntactic and morphological units are identical in all natural languages, but the differences at this "semantic" level are considerably smaller than those at the level of the phenogrammatical syntax and morphology (cf. Curry, l.c.). - In order to exclude the constructional homonymy from the grammar of the IL it is further desirable that its terminal strings should not only contain the lexical and morphological units (arguments), but also the symbols of syntactic relation (functors), [8] so that the derivational history of each string could be inferred from that string in a single way (but cf. §5 as to the order of elements).

4. It is possible to formulate a grammar of such an IL, having not more than ten or so rules of the "context-free" type and a counter; for a first sketch, see Appendix. In these rules the functors are denoted by the symbols R_a (relation of agens

to action), R_p (relation of the object, "patients", to the action), R_d (relation of an adjunct to its nucleus); the symbols of the form Q_j^i denote the coordination or apposition of i members. The generated strings can be regarded as formulae written similarly to the Polish notation, where a variable is a sequence between two "+" signs (or between this sign and the beginning or end of the string) and each symbol of the form R_{index} is a functor with two arguments. (As to the linguistic interpretation, the first of these arguments depends on the second, e.g. $R_a + S + V$ denotes the S - i.e. noun - as the agens of V .)

The morphological units are represented here as "suffixes", standing right of a "-" sign, which connects them with the lexical unit (i.e. with the leftmost symbol of a substring as Subst-pl-indef). The abbreviations used are to be understood roughly as the units known from classical linguistics, but their proper interpretation is, of course, given only through the rules of the grammar; its connection with the units of individual languages is to be completed by the algorithms of analysis (for the input languages) and synthesis (for the output languages). In these algorithms the "syntactic derivation" [9] has to be incorporated, so that verbal nouns are to be translated into the IL as verbal forms, adjectival adverbs (simply, briefly, well etc.) - as forms of adjectives (not as separate words) etc.

In such an IL there will be no difference between an active and a corresponding passive construction, between a verbal clause and a nominalized construction etc. The rules of the choice between such constructions in the output language are to be incorporated in the algorithms of synthesis. (In a good translation this choice does not depend essentially on the construction used in the input text, but rather on grammatical and other rules of the output language itself.)

5. Terminal strings generated by the proposed grammar are units of the language, not parts of texts. The word order does not play any grammatical role in the proposed IL and therefore

is not determined by its grammar; it is not identical with the order of the symbols of the terminal strings. In the process of translation it will be possible to leave the word order, in essential, as it was in the input text, eliminating only the rules of arrangement that are characteristic for the input language. (As to the auxiliary words, they appear in the IL only as "suffixes" or "functors".) The word order may be marked in the text of the IL by special indexes. The structural description of a sentence in the input language is to be restated (by means of the algorithm of the analysis) in such a way that a new structural description in terms of the IL is associated to the sentence, containing only phrase-structure rules; e.g. subordinated clauses (as well as nominalizations) are treated here as members of the main clause. For the synthesis of the output text, context restricted rules and transformations must be formulated in a similar way as in a generative grammar of the output language, accompanied by rules of choice for optional transformations.

6. The practical usefulness of such an IL naturally is not granted, until the algorithms for the respective languages are elaborated and tested; but further work on the grammar of the IL (which must include a study of many languages of different types from the standpoint of the relation of form and function) will be of interest not only for machine translation, but also for a general theory of language.

REFERENCES

- [1] This "short" is, essentially, an abstract from an article published in Czech (P.Sgall, Převodní jazyk a teorie gramatiky, Slovo a slovesnost 24, 1963, 114-128).

- [2] CHOMSKY, N.: Syntactic Structures, 's-Gravenhage 1957; On Certain Formel Properties of Grammars, Inf. and Control 2, 1959, 137-167; On the Notion "Rule of Grammar", Structure of Language and its Mathematical Aspects (PSAM 12), 1961, 6-24; Explanatory Models in Linguistics, Proc. of Intern. Congr. for Logic, Methodology and Philisophy of Science, Stanford Univ., ed. 1962; Y.Bar-Hillel, M.Perles, E.Shamir, On Formal Properties of Simple Phrase Structure Grammars, Jerusalem 1962; Y.Bar-Hillel. The present Status of Automatic Translation of Languages, Advances in Computers I, 1960, 91-163; K.Čulík, Some Axiomatic Systems for Formal Grammars and Languages, Preprint of the Proc. of the IFIP Congress 62, Munich 1962, 134-137.

- [3] As to a program of such a theory, see J.Fodor, J.Katz, The Structure of a Semantic Theory, Language 39, 170-210.

- [4] SÉCHEHAYE, A.: Essai sur la structure logique de la phrase, Paris 1926; Ch.Bally, Linguistique générale et ling. française, Paris 1932; V.Mathesius, Čestina a obecný jazykozpyt, Praha 1947 etc.

- [5] For a discussion of the problems of IL, cf. Y.Bar-Hillel, o.c. in ref. 2, N.D.Andreyev - S.Y.Fitilov, Jazyk-posrednik mašinного perevoda..., in Tezisy soveščanija po matem. lingistike, Leningrad 1959; I.A.Mel'čuk, Nekotoryje voprosy mašinного perevoda za rubežom, Moskva (VINITI) 1961, 18-25, and his K voprosu o grammatičeskom v jazyke-posrednike, Mašinnyj perevod i prikladnaja lingvistika 4, 1960, 25-45.
- [6] KNOWLTON, K.G.: Sentence Parsing with a Self-Organizing Heuristic Program, Cambridge (MIT) 1962, Chapter 1 (mimeogr.); cf. S.Y.Fitilov O modelirovanii sintaksisa v strukturnoj lingvistike, Problemy strukturnoj lingvistiki, Moskva 1962, 100-114.
- [7] DOKULIL, M. - DANĚŠ, Fr.: K tzv. významové a mluvnické stavbě věty, in O vědeckém poznání soudobých jazyků, Praha 1958, 231-246; I.I. Revzin, Formal'nyj i semantičeskij analiz sintaksičeskich svjazej v jazyke, in Priimenenije logiki v nauke i technike, Moskva, 1960, 119-139; H.B.Curry, Some Logical Aspects of Grammatical Structure, PSAM 12, 56-68 (esp. p.65).
- [8] Cf. the model described by V.H.Yngve, A Model and an Hypothesis for Language Structure, Proc. of Am. Philos. Soc. 104, Nr 5, 1960, 444ss.
- [9] KURYLOVIČ, J.: Dérivation lexicale et dérivation syntaxique, Bull. de la Soc.ling. de Paris 37, 1936, 79-92.

APPENDIX

I. The Grammar of an Intermediate Language

- 1 Sent \longrightarrow $\left\{ \begin{array}{l} \text{Enunt} \\ \text{Interr} \\ \text{Voc} \end{array} \right.$
- 2 Enunt \longrightarrow $\left\{ \begin{array}{l} \text{Pred} \\ \text{S} \\ \text{A} \end{array} \right.$ - Modal
- 3 Interr \longrightarrow $\left\{ \begin{array}{l} \text{Pred} \\ \text{S} \\ \text{A} \end{array} \right.$ - interr
- 4 Voc \longrightarrow $\left\{ \begin{array}{l} \text{S} \\ \text{A} \end{array} \right.$ - voc
- 5 Pred \longrightarrow $\left\{ \begin{array}{l} (R_a + NP_1 +) VP \\ R_a + NP_1 + NP_0 - Asp - Mod - Temp \\ A - Asp - Mod - Temp \end{array} \right.$
- 6 VP \longrightarrow $\left\{ \begin{array}{l} R_p + NP_1 + V_{tr} \\ R_p + NP_1 + R_p + NP_1 + V_{bitr} \end{array} \right.$
- 7 λ \longrightarrow $R_d + NP_0 + \lambda$ (where λ stands for
Sent, Enunt, Interr,
Voc, NP_0 , NP_1 , VP)

- 8 $Y \rightarrow \left\{ \begin{array}{l} Q_{cop}^i \\ Q_{grad}^i \\ Q_{disj}^i \\ Q_{adv}^2 \\ Q_{contr}^2 \end{array} \right\} + Y + \dots + Y$ i -times (where i is an integer greater than 1 and Y stands for the same symbols as X above, excepted for Sent)
- 9 $Z \rightarrow Q_{ap}^i + Z + \dots + Z$ i times, with i as above and Z standing for Sent, NP_0 , NP_1 , VP
- 10 $NP_0 \rightarrow \left\{ \begin{array}{l} S - (Det) \\ A - (Det) \\ Pred - Rel - Det \end{array} \right.$
- 11 $NP_1 \rightarrow \left\{ \begin{array}{l} S \\ A \\ Pred - Rel (- indir) \end{array} \right.$ (Note: indir "means" an indirect question)
- 12 $S \rightarrow$ Subst - Num - Def
- 13 $V_{tr} \rightarrow$ Verb_{tr} - (Num - Def -) Asp - Mod - Temp
- 14 $V_{bitr} \rightarrow$ Verb_{bitr} - (Num - Def -) Asp - Mod - Temp
- 15 $VP \rightarrow$ Verb_{intr} - (Num - Def -) Asp - Mod - Temp
- 16 $Pred \rightarrow$ Verb_{impers} - (Num - Def -) Asp - Mod - Temp
- 17 $A \rightarrow \left\{ \begin{array}{l} Adj - (Num - Def -) Comp \\ Adv (- Num - Def) \end{array} \right.$
- 18 $Num \rightarrow$ sg, pl, anum
- 19 $Def \rightarrow$ not, cont, possess, indef
- 20 $Comp \rightarrow$ pos, comp, sup, elat

21	Asp	→	mom, ingr, fin, result, dur, iter, stat
22	Mod	→	ind, opt, deb, vol, possib, perm, hort, facult
23	Temp	→	praes, praet, fut, gnom
24	Rel	→	simult, praec, succ
25	Det	→	gener, instr, caus, cond, modus, loc ₁ - loc ₃₆ , temp ₁ - temp ₇ , ...
26	Modal	→	enunt, desid, imper
27	Subst	→	America, he, Jackson, Mathematician, mathematics, problem, physicist, theorem, student, year.
28	Adj	→	famous, gifted, important, other succesful, this young
29	Adv	→	yesterday, today
30	Verb _{impres}	→	rain
31	Verb _{intr}	→	try
32	Verb _{tr}	→	know, prove, solve
33	Verb _{bitr}	→	teach

Note:

The grammar has the form of a context-free phrase structure grammar, with a counter, using some conventions known from the work of Chomsky; its initial symbol is Sent. There are several open questions, as the representation of pronouns; the list of symbols is not complete (namely in rule 25, which should include a number of adverbial and attributive functions, such as purpose, quality etc.).

The English words given for the sake of briefness in the rules 27-33 as well as in the examples of §II, symbolize the respective units of the IL (the semoglyphs).

II. Examples of the derivation
of sentences in the IL
(by S.Komrsková)

In the examples the word-order of the text is not denoted in the IL-form of the derived sentences. The derivation of most sentences here begins by the application of rule 1a, which is not stated explicitly (for the sake of brevity); the application of other rules is given left of each derived line.

Some symbols in our examples do not agree exactly with the forms given in the sketch of grammar (§I); in §II they are written in agreement with Czech orthographical usage, as:

Enunc	=	Enunt	
kont	=	cont	(i.e., definite by context, anaphorical)
noc	=	not	(i.e., notional definiteness)
prec	=	praec	(i.e., praecedens)
suko	=	succ	(i.e., succedens)
etc.			

In rule 25, "gener" means an unspecified relation (as in "the top of the body", "green paper"), loc, -loc₃₆ are the different local (adverbial) relations as in, int₉, on, outs, from, ..., and temp₁ - temp₇ are temporal relations including since, till specified by the tables of "Entwurf der Grammatik einer Interlingua", p.3. In other respects, the "Entwurf" as an instruction for generating IL-sentences is to be replaced now by the present paper.

1)

Enunc

- 2a Pred - Modal
- 5a $R_a + NP_1 + VP - \text{Modal}$
- 6a $R_a + NP_1 + R_p + NP_1 + V_{tr} - \text{Modal}$
- 11a, 11a $R_a + S + R_p + S + V_{tr} - \text{Modal}$
- 12-17 $R_a + \text{Subst} - \text{Num} - \text{Def} + R_p + \text{Subst} - \text{Num} - \text{Def} +$
 $+ \text{Verb}_{tr} - \text{Asp} - \text{Mod} - \text{Temp} - \text{Modal}$
- 18-33 $R_a + \text{mathematician} - \text{sg} - \text{kont} + R_p + \text{theorem} -$
 $- \text{sg} - \text{indef} + \text{prove} - \text{dur} - \text{ind} - \text{gnóm} - \text{enunc}$

E.g.: The mathematician proves a theorem.

Matematik dokazyvajet teoremu.

2)

Enunc

- 2a Pred - Modal
- 5a $R_a + NP_1 + VP - \text{Modal}$
- 6a $R_a + NP_1 + R_p + NP_1 + V_{tr} - \text{Modal}$
- 11a, 11a $R_a + S + R_p + S + V_{tr} - \text{Modal}$
- 12 - 17 $R_a + \text{Subst} - \text{Num} - \text{Def} + R_p + \text{Subst} - \text{Num} - \text{Def} -$
 $- + \text{Verb}_{tr} - \text{Asp} - \text{Mod} - \text{Temp} - \text{Modal}$
- 18-33 $R_a + \text{mathematician} - \text{pl} - \text{indef} + R_p + \text{theorem} -$
 $- \text{sg} - \text{kont} + \text{prove} - \text{result} - \text{ind} - \text{fut} - \text{enunc}$

E.g.: Mathematicians will prove the theorem.

Teoremu dokazut matematiki.

3)

Enunc

2a Pred - Modal

5a₁ R_a + NP₁ + VP - Modal

6a R_a + NP₁ + R_p + NP₁ + V_{tr} - Modal

11a, 11a R_a + S + R_p + S + V_{tr} - Modal

12-17 R_a + Subst - Num - Def + R_p + Subst - Num - Def +
+ V_{tr} - neg - Asp - Mod - Temp - Modal

18-33 R_a + mathematician - sg - kont + R_p + theorem -
- sg - indef + prove - neg - result - ind - fut -
- enunc

E.g.: The mathematician will not prove a theorem.

Matematik teoremy n'e dokazet.

4)

Enunc

2a Pred - Modal

5a₁ R_a + NP₁ + VP - Modal

6a R_a + NP₁ + R_p + V_{tr} - Modal

11a, 11a R_a + S + R_p + S + V_{tr} - Modal

12-17 R_a + Subst - Num - Def + P_p + Subst - Num - Def +
+ Verb_{tr} - Asp - Mod - Temp - Modal

18-33 R_a + mathematician - sg - kont + R_p + theorem -
- sg - indef + prove - result - ind - pret - enunc

E.g.: The mathematician proved a theorem.

Matematik dokazal teoremu.

- 5) Enunc
- 2a Pred - Modal
- 5a₁ R_a + NP₁ + VP - Modal
- 6a R_a + NP₁ + R_p + NP₁ + V_{tr} - Modal
- 11a, 11a R_a + S + R_p + V_{tr} - Modal
- 12-17 R_a + Subst - Num - Def + R_p + Subst - Num - Def +
+ Verb_{tr} - neg - Asp - Mod - Temp - Modal
- 18-33 R_a + mathematician - sg - kont + R_p + theorem -
pl - indef + prove - neg - result - ind - prêt -
- enunc

E.g.: The mathematician didn't prove theorems.
Matematik ne dokazal [✓]teoremu.

- 6) Enunc
- 2a Pred - Modal
- 5a₁ R_a + NP₁ + VP - Modal
- 6a R_a + NP₁ + R_p + NP₁ + V_{tr} - Modal
- 7 R_a + NP₁ + R_p + R_d + NP_o + NP₁ + V_{tr} - Modal
- 11a, 10a,
11a R_a + S + R_p + R_d + S + S + V_{tr} - Modal
- 12-17 R_a + Subst - Num - Def + R_p + R_d + Subst - Num -
- Def + Subst - Num - Def + Verb_{tr} - Asp - Mod -
- Temp - Modal
- 18-33 R_a + mathematician - sg - kont + R_p + R_d + he -
- pl - noc + theorem - sg - poses + prove -
- result - ind - prêt - enunc

E.g.: The mathematician proved their theorem.

Matematik dokazal ich t^voremu.

7)

Enunc

2a Pred - Modal

5b $R_a + NP_1 + NP_0 - asp - Mod - Temp - Modal$

8a $R_a + NP_1 + Q_{kop}^2 + NP_0 + NP_0 : - Asp - Mod - Temp -$
- Modal

11a, 10a,

10a $R_a + S + O^2 + S + S : - Asp - Mod - Temp - Modal$

12-17 $R_a + Subst - Num - Def + Q_{kop}^2 + Subst - Num - Def +$
+ Subst - Num - Def : - Asp - Mod - Temp - Modal

18-33 $R_a + \underline{he} - sg - noc + Q_{kop}^2 + \underline{mathematician} - sg -$
- indef + physicist - sg - indef : - dur - ind -
- préz - enunc

E.g.: He is a mathematician and physicist.

On matematik i fizik.

8)

Enunc

2a Pred - Modal

5a $R_a + NP_1 + VP - Modal$

7, 6a $R_a + R_d + NP_0 + NP_1 + R_p + NP_1 + V_{tr} - Modal$

7 $R_a + R_d + NP_0 + NP_1 + R_p + R_d + NP_0 + NP_1 + V_{tr} - Modal$

10a, 11a

10a, 11a $R_a + R_d + S + S + R_p + R_d + S + S + V_{tr} - Modal$

12-17 $R_a + R_d + Subst - Num - Def + Subst - Num - Def +$
+ $R_p + R_d + Subst - Num - Def + Subst - Num - Def +$
+ Verb_{tr} - Asp - Mod - Temp - Modal

18-33 $R_a + R_d + \text{America} - \text{sg} - \text{noc} + \text{mathematician} -$
 $- \text{sg} - \text{kont} + R_p + R_d + \text{he} - \text{pl} - \text{noc} - + \text{theorem} -$
 $- \text{sg} - \text{poses} + \text{prove} - \text{result} \text{ ind} - \text{prét} - \text{enunc}$

E.g.: The American mathematician proved their theorem.
 Amerikanskij matematik dokazal ich tēoremu.

9)

Enunc

2a Pred - Modal

5a $R_a + NP_1 + VP - \text{Modal}$

7, 6a $R_a + R_d + NP_0 + NP_1 + R_p + NP_1 + V_{tr} - \text{Modal}$

8a, 7 $R_a + R_d + Q_{kop}^2 + NP_0 : + NP_1 + R_p + R_d + NP_0 +$
 $+ NP_1 + V_{tr} - \text{Modal}$

10b, 10b, 11a,

10a, 11a $R_a + R_d + Q_{kop}^2 + A + A : + S + R_p + R_d + S + S +$
 $+ V_{tr} - \text{Modal}$

12-17 $R_a + R_d + Q_{kop}^2 + \text{Adj} - \text{komp} + \text{Adj} - \text{komp} : + \text{Subst} -$
 $- \text{Num} - \text{Def} + R_p + R_d + \text{Subst} - \text{Num} - \text{Def} + \text{Subst} -$
 $- \text{Num} - \text{Def} + \text{Verb}_{tr} - \text{Asp} - \text{Mod} - \text{Temp} - \text{Modal}$

18-33 $R_a + R_d + Q_{kop}^2 + \text{young} - \text{poz} + \text{gifted} - \text{poz} : +$
 $+ \text{mathematician} - \text{sg} - \text{kont} + R_p + R_d + \text{he} - \text{pl} -$
 $\text{noc} + \text{theorem} - \text{sg} - \text{poses} + \text{prove} - \text{result} - \text{ind} -$
 $- \text{prét} - \text{enunc}$

E.g.: The young and gifted mathematician proved their theorem.

Molodoj i talantlivyj matematik dokazal ich tēoremu.

10)

Enunc

2a Pred-Modal

5a₁ R_a + NP₁ + VP - Modal

9, 6a R_a + Q_{ap}² + NP₁ + NP₁ : + R_p + NP₁ + V_{tr}

7,7 R_a + Q_{ap}² + NP₁ + R_d + NP₀ + NP₁ : + R_p + R_d +
+ NP₀ + NP₁ + V_{tr} - Modal

11a, 10a,

11a,10a, R_a + Q_{ap}² + S + R_d + S + S : + R_p + R_d + S + S +

11a + V_{tr} - Modal

12-17 R_a + Q_{ap}² - Subst - Num - Def + R_d + Subst - Num -
- Def + Subst - Num - Def : + R_p + R_d + Subst -
- Num - Def + Subst - Num - Def + Verb_{tr} - Asp -
- Mod - Temp - Modal

18-33 R_a + Q_{ap}² + Jackson - sg - noc + R_d + America - sg -
- noc + mathematician - sg - indef : + R_p + R_d +
+ he - pl - noc + theorem - sg - poses + prove -
result - ind - prét - enunc

E.g.: Jackson, an American mathematician, proved their
theorem.

Džekson, amerikánskiĵ matematik, dokazal ich tĕo-
remu.

11)

Enunc

2a Pred - Modal

5a₁ R_a + NP₁ + VP + Modal

7 R_a + NP₁ + R_d + NP₀ + VP - Modal

- 7, 6a $R_a + R_d + NP_0 + NP_1 + R_d + NP_0 + R_p + NP_1 + V_{tr} -$
- Modal
- 7, 7 $R_a + R_d + NP_0 + R_d + NP_0 + NP_1 + R_d + NP_0 + R_p +$
 $+ R_d + NP_0 + NP_1 + V_{tr} -$ Modal
- 10b, 10a, 11a $R_a + R_d + A + R_d + S + S + R_d + S -$ Def + $R_p +$
10a, 10a, 11a $+ R_d + S + S + V_{tr} -$ Modal
- 12-17 $R_a + R_d + Adj - komp + R_d + Subst - Num - Def +$
 $+ Subst - Num - Def + R_d + Subst - Num - Def -$
 $- Det + R_p + R_d + Subst - Num - Def + Subst - Num -$
 $- Def + Verb_{tr} - Asp - Mod - Temp -$ Modal
- 18-33 $R_a + R_d + \text{young} - poz + R_d + \text{America} - sg - noc +$
 $+ \text{mathematicien} - sg - kont + R_d + \text{year} - sg - noc -$
 $- temp_i + R_p + R_d + \text{he} - pl - noc + \text{theorem} - sg -$
 $- poses + \text{prove} - result - ind - prêt - enunc$

E.g.: The young American mathematician proved their
theorem a year ago.

Molodoj amerikanskij matematik dokazal ich téoremu
god tomu nazad.

12)

Enunc

2a Pred - Modal

5a₁ $R_a + NP_1 + VP -$ Modal

7 $R_a + NP_1 + R_d + NP_0 + VP -$ Modal

7, 7 $R_a + R_d + NP_0 + NP_1 + R_d + NP_0 + R_d + NP_0 + VP -$
Modal

- 7, 6 $R_a + R_d + NP_0 + R_d + NP_0 + NP_1 + R_d + NP_0 + R_d +$
 $+ NP_0 + R_p + NP_1 + V_{tr} - \text{Modal}$
- 7 $R_a + R_d + NP_0 + R_d + NP_0 + NP_1 + R_d + NP_0 + R_d +$
 $+ NP_0 + R_p + R_d + NP_0 + NP_1 + V_{tr} - \text{Modal}$
- 10b, 10a, 11a,
 10b₂, 10b₂, $R_a + R_d - A + R_d + S + S + R_d + A - \text{Det} + R_d +$
 10a, 10a $+ A - \text{Det} + R_p + R_d + A + S + V_{tr} - \text{Modal}$
- 12-17 $R_a + R_d + \text{Adj} - \text{komp} + R_d + \text{Subst} - \text{Num} - \text{Def} +$
 $+ \text{Subst} - \text{Num} - \text{Def} + R_d + \text{Adj} - \text{komp} - \text{Det} + R_d +$
 $+ \text{Adv} - \text{Det} + R_p + R_d + \text{Subst} - \text{Num} - \text{Def} + \text{Subst} -$
 $- \text{Num} - \text{Def} + \text{Verb}_{tr} - \text{Asp} - \text{Mod} - \text{Temp} - \text{Modal}$
- 18-33 $R_a + R_d + \text{Young} - \text{poz} + R_d + \text{America} - \text{sg} - \text{noc} +$
 $+ \text{mathematician} - \text{sg} - \text{kont} + R_d + \text{successful} -$
 $- \text{poz} - \text{modus} + R_d + \text{yesterday} - \text{temp}_1 + R_p + R_d +$
 $+ \text{he} - \text{pl} - \text{noc} - \text{theorem} - \text{sg} - \text{poses} + \text{prove} -$
 $- \text{result} - \text{ind} - \text{prét} - \text{enunc}$

E.g.: The young American mathematician proved succesfully
 their theorem yesterday.
 Molodoj amerikanskij matěmatik uspěsno dokazal ich
 těoremu včera.

- 13) Enunc
- 8d $Q_{adv}^2 + \text{Enunc} + \text{Enunc} :$
- 2a, 2a $Q_{adv}^2 + \text{Pred} - \text{Modal} + \text{Pred} - \text{Modal} :$
- 5a, 5a $Q_{adv}^2 + R_a + NP_1 + VP - \text{Modal} + R_a + NP_1 + VP -$
 $- \text{Modal} :$
- 6a, 6a $Q_{adv}^2 + R_a + NP_1 + V_{intr} - \text{Modal} + R_a + NP_1 + R_p +$
 $+ NP_1 + V_{tr} - \text{Modal} :$

- 7 $Q_{adv}^2 + R_a + NP_1 + V_{intr} - \text{Modal} + R_a + NP_1 + R_p +$
 $+ R_d + NP_0 + NP_1 + V_{tr} - \text{Modal}$
- 11a, 11a,
 10a, 11a, $Q_{adv}^2 + R_a + S + V_{intr} - \text{Modal} + R_a + S + R_p + R_d +$
 $+ S + S + V_{tr} - \text{Modal} :$
- 12-17 $Q_{adv}^2 + R_a + \text{Subst} - \text{Num} - \text{Def} + \text{Verb}_{intr} - \text{Asp} -$
 $- \text{Mod} - \text{Temp} - \text{Modal} + R_a + \text{Subst} - \text{Num} - \text{Def} +$
 $+ R_p + R_d + \text{Subst} - \text{Num} - \text{Def} + \text{Subst} - \text{Num} -$
 $- \text{Def} + \text{Verb}_{tr} - \text{neg} - \text{Asp} - \text{Mod} - \text{Temp} - \text{Modal} :$
- 18-33 $Q_{adv}^2 + R_a + \underline{I} - \text{sg} - \text{noc} + \underline{\text{reflect}} - \text{dur} - \text{ind} -$
 $- \text{prét} - \text{enunc} + R_a + \underline{I} - \text{sg} - \text{noc} + R_p + R_d +$
 $+ \underline{\text{he}} - \text{pl} - \text{noc} + \underline{\text{theorem}} - \text{sg} - \text{poses} + \underline{\text{prove}} -$
 $- \text{neg} - \text{result} - \text{ind} - \text{prét} - \text{enunc} :$

E.g.: I reflected but I didn't prove their theorem.
 Ja razmýšľal(a) no ich tēoremu ja ne dokazal(a).

- 14) Enunc
- 8c $Q_{disj}^2 + \text{Enunc} + \text{Enunc}$
- 2a, 2a $Q_{disj}^2 + \text{Pred} - \text{Modal} + \text{Pred} - \text{Modal} :$
- 5a, 5a $Q_{disj}^2 + R_a + NP_1 + VP - \text{Modal} + R_a + NP_1 + VP -$
 $- \text{Modal} :$
- 6a, 6a $Q_{disj}^2 + R_a + NP_1 + R_p + NP_1 + V_{tr} - \text{Modal} + R_a +$
 $+ NP_1 + R_p + NP_1 + V_{tr} - \text{Modal} :$
- 7, 7 $Q_{disj}^2 + R_a + NP_1 + R_p + R_d + NP_0 + NP_1 + V_{tr} -$
 $- \text{Modal} + R_a + NP_1 + R_p + R_d + NP_0 + NP_1 + V_{tr} -$
 $- \text{Modal} :$

11a,10b,11a, $Q^2_{disj} + R_a + S + R_p + R_d + A + S + V_{tr} -$
 11a,10a,11a $- Modal + R_a + S + R_p + R_d + S + S + V_{tr} - Modal:$
 12-17 $Q^2_{disj} + R_a + Subst - Num - Def + R_p + R_d - Adj -$
 $- komp + Subst - Num - Def + Verb_{tr} - Asp - Mod -$
 $- Temp - Modal + R_a + Subst - Num - Def + R_p +$
 $+ R_d + Subst - Num - Def + Subst - Num - Def +$
 $+ Verb_{tr} - Asp - Mod - Temp - Modal:$
 18-33 $Q^2_{disj} + R_a + \underline{I} - pl - noc + R_p + R_d + \underline{this} - poz +$
 $+ \underline{Problem} - sg - kont + \underline{solve} - result - ind -$
 $- fut - enunc + R_a + I - pl - noc + R_p + R_d + \underline{he} -$
 $- pl - noc + \underline{theorem} - sg - poses + \underline{prove} -$
 $- result - ind - fut - enunc :$

E.g.: We shall either solve this problem or we shall
 prove their theorem.
 My ili razrešim etu problemu, ili dokažem ich
 teoremu.

15)

Enunc

2a Pred - Modal

5a $R_a + NP_1 + VP - Modal$

6a $R_a + NP_1 + R_p + NP_1 + V_{tr} - Modal$

11c $R_a + NP_1 + R_p + Pred - Rel + V_{tr} - Modal$

5a $R_a + NP_1 + R_p + R_a + NP_1 + VP - Rel + V_{tr} -$
 $- Modal$

6a $R_a + NP_1 + R_p + R_a + NP_1 + R_p + NP_1 + V_{tr} - Rel +$
 $+ V_{tr} - Modal$

- 7 $R_a + NP_1 + R_p + R_a + NP_1 + R_p + R_d + NP_0 + NP_1 +$
 $+ V_{tr} - Rel + V_{tr} - Modal$
- 11a, 11a,
 10a, 11a, $R_a + S + R_p + R_a + S + R_p + R_d + S + S + V_{tr} -$
 $R_{el} + V_{tr} - Modal$
- 12-17 $R_a + Subst - Num - Def + R_p + R_a + Subst - Num -$
 $- Def + R_p + R_d + Subst - Num - Def + Subst -$
 $- Num - Def + Verb_{tr} - Asp - Mod - Temp - Rel +$
 $+ Verb_{tr} - Asp - Mod - Temp - Modal$
- 18-33 $R_a + \underline{I} - sg - noc + R_p + R_a + Jackson - sg - noc +$
 $+ R_p + R_d + \underline{he} - pl - noc + \underline{theorem} - sg - poses +$
 $+ \underline{prove} - result - ind - \acute{p}r\acute{e}t - prec + \underline{Know} -$
 $- fin - ind - \acute{p}r\acute{e}t - enunc$

E.g.: I knew Jackson had proved their theorem.
 Ja uznal(a), cto Džekson dokazal ich téoremu.

16)

Enunc

- 2a Pred - Modal
- 5b $R_a + NP_1 + NP_0 - Asp - Mod - Temp - Modal$
- 7 $R_a + R_d + NP_0 + NP_1 + NP_0 - Asp - Mod - Temp -$
 $- Modal$
- 10c $R_a + R_d + Pred - Rel - Det + NP_1 + NP_0 - Asp -$
 $- Mod - Temp - Modal$
- 5a $R_a + R_d + R_a + NP_1 + VP - Rel - Det + NP_1 +$
 $+ NP_0 - Asp - Mod - Temp - Modal$
- 6a $R_a + R_d + R_a + NP_1 + R_p + NP_1 + V_{tr} - Rel - Det +$
 $+ NP_1 + NP_0 - Asp - Mod - Temp - Modal$

- 11a, 11a,
11a, 10b,
12-17
18-33
- $R_a + R_d + R_a + S + R_p + S + V_{tr} - Rel - Det -$
 $- S + S - Asp - Mod - Temp - Modal$
- $R_a + R_d + R_a + Subst - Num - Def + R_p + Subst -$
 $- Num - Def + Verb_{tr} - Asp - Mod - Temp - Rel -$
 $- Det + Subst - Num - Def + Adj - komp - Asp -$
 $Mod - Temp - Modal$
- $R_a + R_d + R_a + \underline{Jackson} - sg - noc + R_p + \underline{theorem} -$
 $- sg - kont + \underline{prove} - result - ind - \underline{prét} - prec -$
 $- atr + \underline{theorem} - sg - kont + \underline{important} - poz -$
 $- dur - ind - gnóm - enunc$

E.g.: The theorem that Jackson proved is important.
 (The theorem proved by Jackson is important.)
 Teorema, ktoruju dokazal Džekson, važnaja.
 (Dokazannaja Džolsonom teorema važnaja.)

17)

Enunc

- 2a Pred - Modal
- 5a $R_a + NP_1 + VP - Modal$
- 7 $R_a + NP_1 + R_d + NP_0 + VP - Modal$
- 10c, 6a $R_a + NP_1 + R_d + Pred - Rel - Det + R_p + NP_1 +$
 $+ V_{tr} - Modal$
- 5a, 7 $R_a + NP_1 + R_d + R_a + NP_1 + VP - Rel - Det + R_d +$
 $+ NP_0 + NP_1 + V_{tr} - Modal$
- 6a $R_a + NP_1 + R_d + R_a + NP_1 + R_p + NP_1 + V_{tr} - Rel -$
 $- Det + R_p + R_d + NP_0 + R_d + NP_0 + NP_1 + V_{tr} -$
 $- Modal$

11a,11a,11a,
 10b,10b,11a, $R_a + S + R_d + R_a + S + R_p + S + V_{tr} - Rel -$
 $- Det + R_p + R_d + A + R_d + A + S + V_{tr} - Modal$
 12-17 $R_a + Subst - Num - Def + R_d + R_a + Subst -$
 $- Num - Def + R_p + Subst - Num - Def + Verb_{tr} -$
 $- Asp - Mod - Temp - Rel - Det + R_p + R_d + Adj -$
 $- komp + R_d + Adj - komp + Subst - Num - Def +$
 $+ Verb_{tr} - Asp - Mod - Temp - Modal$
 18-33 $R_a + \underline{Jackson} - sg - noc - R_d + R_a + \underline{Jackson} -$
 $- sg - noc + R_p + \underline{theorem} - sg - kont + prove -$
 $- result - ind - pr\acute{e}t - prec - temp_i + R_p + R_d +$
 $+ \underline{other} - poz + R_d + \underline{important} - poz + \underline{problem} -$
 $- sg - indef + \underline{solve} - ingr - ind - pr\acute{e}t -$
 $- enunc$

E.g.: When Jackson had proved the theorem, he started
 to solve another important problem.

(After having proved the theorem, Jackson
 started to solve another important problem.)

Kogda Džekson dokazal teoremu, on stal rešit'
 druguju važnuju problemu.

(Džekson, dokazav teoremu, stal rešit' druguju
 važnuju problemu.)

18)

Enunc

2a

Pred - Modal

5a

$R_a + NP_1 + VP - Modal$

7

$R_a + NP_1 + R_d + NP_0 + VP - Modal$

10c, 6a $R_a + NP_1 + R_d + Pred - Rel - Det + R_p + NP_1 + V_{tr} -$
 $- Modal$

$R_a + NP_1 + R_d + VP - Rel - Det + R_p + R_d + NP_0 +$
 $+ NP_1 + V_{tr} - Modal$

$R_a + NP_1 + R_d + R_p + NP_1 + V_{tr} - Rel - Det + R_p +$
 $+ R_d + NP_0 + R_d + NP_0 + NP_1 + V_{tr} - Modal$

$R_a + S + R_d + R_p + S + V_{tr} - Rel - Det + R_p +$
 $+ R_d + A + P_d + A + S + V_{tr} - Modal$

12-17 $R_a + Subst - Num - Def + R_d + R_p + Subst - Num -$
 $- Def + Verb_{tr} - Asp - Mod - Temp - Rel - Det +$
 $R_p + R_d + Adj - komp + R_d + Adj - komp + Subst -$
 $- Num - Def + Verb_{tr} - Asp - Mod - Temp - Modal$

18-33 $R_a + \underline{Jackson} - sg - noc + R_d + R_p + \underline{theorem} -$
 $- sg - kont + \underline{prove} - result - ind - \acute{p}r\acute{e}t - prec -$
 $- temp_i + R_p + R_d + \underline{other} - poz + R_d + \underline{important} -$
 $- poz + \underline{problem} - sg - indef + \underline{solve} - ingr -$
 $- ind - \acute{p}r\acute{e}t - enunc$

E.g.: When the theorem was proved, Jackson started to
 solve another important problem.

(After the theorem had been proved, Jackson
 started to solve another important problem.)

Kogda byla teorema dokazana, stal Džekson
 rešit' važnuju druguju problemu.

(Posle dokazanija teoremy stal Džekson resit'
 važnuju druguju problemu.)

19)

Enunc

2a Pred - Modal

5a VP - Modal

7 $R_d + NP_0 + VP - \text{Modal}$

10c, 6a $R_d + \text{Pred} - \text{Rel} - \text{Det} + R_p + NP_1 + V_{tr} - \text{Modal}$

5a $R_d + R_a + NP_1 + VP - \text{Rel} - \text{Det} + R_p + NP_1 +$
 $+ V_{tr} - \text{Modal}$

6a $R_d + R_a + NP_1 + R_p + NP_1 + V_{tr} - \text{Rel} - \text{Det} + R_p +$
 $+ NP_1 + V_{tr} - \text{Modal}$

7 $R_d + R_a + NP_1 + R_p + R_d + NP_0 + NP_1 + V_{tr} - \text{Rel} -$
 $- \text{Det} + R_p + NP_1 + V_{tr} - \text{Modal}$

11a, 10b,
 11a, 11a, $R_d + R_a + S + R_p + R_d + A + S + V_{tr} - \text{Rel} - \text{Det} +$
 $+ R_p + S + V_{tr} - \text{Modal}$

12-17 $R_d + R_a + \text{Subst} - \text{Num} - \text{Def} + R_p + R_d + \text{Adj} -$
 $- \text{komp} + \text{Subst} - \text{Num} - \text{Def} + \text{Verb}_{tr} - \text{Asp} - \text{Mod} -$
 $- \text{Temp} - \text{Rel} - \text{Det} + R_p + \text{Subst} - \text{Num} - \text{Def} +$
 $+ \text{Verb}_{tr} - \text{Asp} - \text{Mod} - \text{Temp} - \text{Modal}$

18-33 $R_d + R_a + I - \text{pl} - \text{noc} + R_p + R_d + \text{this} - \text{poz} +$
 $+ \text{theorem} - \text{sg} - \text{kont} + \text{prove} - \text{result} - \text{ind} -$
 $- \text{fut} - \text{prec} - \text{kond} + R_p + \text{problem} - \text{sg} - \text{kont} +$
 $+ \text{solve} - \text{result} - \text{ind} - \text{fut} - \text{enunc}$

E.g.: If we prove this theorem, the problem will be
 solved.

Jesli my doka[✓]žem etu teoremu, to problema bud'et
 pazre[✓]šena.

20)

Enunc

- 7 $R_d + NP_0 + \text{Enunc}$
 2a $R_d + NP_0 + \text{Pred} - \text{Modal}$
 10b $R_d / + A + \text{Pred} - \text{Modal}$
 12-17 $R_d + \text{Adv} + \text{Verb}_{\text{impers}} - \text{Asp} - \text{Mod} - \text{Temp} - \text{Modal}$
 18-33 $R_d + \underline{\text{today}} + \underline{\text{rain}} - \text{dur} - \text{ind} - \text{préz} - \text{enunc}$

E.g.: It rains today.
 (Segodn'a idet dožd'.)

21)

Enunc

- 3a $\text{Pred} - \text{interr}$
 5a $\text{VP} - \text{interr}$
 6a $R_p + NP_1 + V_{\text{tr}} - \text{interr}$
 7 $R_p + R_d + NP_0 + NP_1 + V_{\text{tr}} - \text{interr}$
 10b, 11a $R_p + R_d + A + S + V_{\text{tr}} - \text{interr}$
 12-17 $R_p + R_d + \text{Adj} - \text{komp} + \text{Subst} - \text{Num} - \text{Def} + \text{Verb}_{\text{tr}} -$
 $- \text{Asp} - \text{Mod} - \text{Temp} - \text{interr}$
 18-33 $R_p + R_d + \underline{\text{this}} - \text{poz} + \underline{\text{theorem}} - \text{sg} - \text{kont} +$
 $+ \underline{\text{prove}} - \text{result} - \text{ind} - \text{préz} - \text{interr}$

E.g.: Has this theorem been proved?
 Eta teorema dokazana?

22)

Sent

- 9 $Q_{\text{ap}}^2 + \text{Sent} + \text{Sent} :$
 1c, 1a $Q_{\text{ap}}^2 + \text{Vok} + \text{Enunc} :$
 4a, 2a $Q_{\text{ap}}^2 + S - \text{vok} + \text{Pred} - \text{Modal} :$

- 5a $Q_{ap}^2 + S - \text{vok} + R_a + NP_1 + VP - \text{Modal}$
- 6a $Q_{ap}^2 + S - \text{vok} + R_a + NP_1 + R_p + NP_1 + V_{tr} -$
- Modal :
- 7 $Q_{ap}^2 + S - \text{vok} + R_a + NP_1 + R_p + R_d + NP_0 + NP_1 +$
+ $V_{tr} - \text{Modal} :$
- 11a, 10b, 11a $Q_{ap}^2 + S - \text{vok} + R_a + S + R_p + R_d + A + S + V_{tr} -$
- Modal :
- 12-17 $Q_{ap}^2 + \text{Subst} - \text{Num} - \text{Def} - \text{vok} + R_a + \text{Subst} -$
- Num - Def + $R_p + R_d + \text{Adj} - \text{komp} + \text{Subst} -$
- Num - Def + $\text{Verb}_{tr} - \text{Asp} - \text{Mod} - \text{Temp} - \text{Modal} :$
- 18-33 $Q_{ap}^2 + \underline{\text{Jackson}} - \text{sg} - \text{noc} - \text{vok} + R_a + \underline{\text{Jackson}} -$
- sg - noc + $R_p + R_d + \underline{\text{this}} - \text{poz} + \text{theorem} -$
- sg - kont + $\underline{\text{prove}} - \text{result} - \text{vol} - \text{fut} -$
- imper :

E.g.: Jackson, prove this theorem!
Džekson, dokaži(te) etu teoremu!

23)

Enunc

- 2a Pred - Modal
- 5b $R_a + NP_1 + VP - \text{Modal}$
- 6h $R_a + NP_1 + R_p + NP_1 + V_{bitr} - \text{Modal}$
- 11a, 11a, 11a $R_a + S + R_p + S + R_p + S + V_{bitr} - \text{Modal}$
- 12-17 $R_a + \text{Subst} - \text{Num} - \text{Def} + R_p + \text{Subst} - \text{Num} -$
- Def + $R_p + \text{Subst} - \text{Num} - \text{Def} + \text{Verb}_{bitr} -$
- Asp - Mod - Temp - Modal

18-33

$R_a + \underline{\text{Jackson}} - \text{sg} - \text{noc} + R_p + \underline{\text{mathematics}} -$
 $- \text{sg} - \text{indef} + R_p + \underline{\text{student}} - \text{pl} - \text{indef} +$
 $+ \underline{\text{teach}} - \text{dur} - \text{ind} - \text{gnóm} - \text{enunc}$

E.g.:

Jackson teaches students mathematics.

Džekson učít studentov matematike.

ВЗАИМОСВЯЗЬ ЭЛЕМЕНТОВ ОТНОШЕНИЯ РУССКИХ И ВЕНГЕРСКИХ ИМЕН

И.Агоч и Д.Варга

Цель настоящей работы состоит в том, чтобы непосредственно сравнивать в суженной территории морфологии два языка, не связанных никаким языковым родством. Все-таки имеется реальная основа к сравнению, так как именно в данной области исследования русский и венгерский языки обладают существенным общим свойством: оба языка имеют чрезвычайно богатую морфологию. Этим свойством и объясняется то, что исследование производится без учёта языков в целом, в виду того, что информации, выраженные морфологическими средствами одного из языков, выражаются, как правило, подобными средствами и в случае другого языка. Но для осуществления подобного исследования мы нуждаемся в расширении круга исследуемых явлений, в целях достижения сравнительно закрытой их системы: кроме именных словоформ мы должны учитывать до известной степени и предлоги в русском языке, выражающие взаимоотношения имен, и последологи в венгерском, играющие роль, подобную прилепам.

Это сопоставление элементов отношений имен может быть полезным с двух точек зрения. С одной стороны, непосредственный бинарный перевод легче осуществим с помощью подобной системы соответствия, хотя бы в тех случаях, когда можно устанавливать связь имени с именем. С другой стороны, сопоставление принесет и теоретическую пользу именно из-за различия характера данных языков. Дело в том, что языковые явления, остающиеся скрытыми при бинарном сопоставлении языков, близких по структуре, если оба языка пользуются такими же неявными средствами, — получают гораздо более четкое, явное выражение при сопоставлении языков с сильно расходящимися структурами. Следовательно, бинарное сравнение языков различной структуры полезно и с точки зрения выработки категорий языка-посредника, более, чем сравнение родственных языков. Особенно поучительным обещает быть срав-

нение с венгерским языком, так как венгерский язык обладает целым рядом таких категорий, которые дают возможность, с одной стороны, отображать особенности отдельных языков на один общий язык (см. в этом отношении высказывания И.А.Мельчука о венгерском языке [1]), с другой стороны, выражать категории какого-то языка, нагруженные несколькими функциями, в более развернутой форме, отдельными морфологическими признаками. В частности, говоря о морфологии, именно в области склонения имен венгерский язык владеет такой богатой системой выражения, которую, кажется, можно считать единственной в своем роде. Это богатство понимается не в смысле формального богатства, дело не в том, что одна функция выражается разными способами, а в том, что венгерский язык обладает чрезвычайно богатой шкалой функций.

В венгерском языке имеется такое невероятно большое число склоняемых форм имен, что много лингвистов прямо отрицали существование категорий падежа в венгерском языке. См. например у Й.Балашша: "В венгерском языке ни падежей, ни склонения, ни предлогов нет; их функцию в предложении выполняют... элементы речи, присоединяемые к именам" [2]. В самом деле, может показаться странным для человека, знающего индоевропейские языки, что в то время, как в русском языке имеется 6, в латынском 5 (или 6), в немецком 4, в французском 3 падежа - в венгерском языке существует по мнению одних языковедов 17 падежей, по мнению других 22, или даже 28 падежей, [3]. Мы хотели избежать того, чтобы ограничение числа падежей в какой-то мере тормозило бы установлению взаимосвязи русских и венгерских падежей, поэтому имели в виду самый богатый список с 28-ю падежами за исключением одного, только формально приведенного падежа^{х/}, который явно отсутствует в венгерском языке.

х/ Это - самостоятельный родительный падеж, функцию которого выполняет именительный или дательный падеж вместе с притяжательным окончанием имени, управляющего словом в имени-тельно-родительном или дательно-родительном падеже.

Падежи эти следующие^{xx/} [4]:

		<u>Окончание:</u>
1. Nominativus	ember человек	Ø
2. Accusativus	embert человека (вижу)	t/at/ot/et/öt
3. Dativus	embernek человеку	nak/nék
4. Instrumentalis	emberrel с человеком	val/vel
5. Illativus	emberbe в человека	ba/be
6. Inessivus	emberben в человеке	ban/ben
7. Elativus	emberről из человека	ból/ből
8. Sublativus	emberre на человека	ra/re
9. Superessivus	emberen на человеке	n/on/en/ön
10. Delativus	emberről с человека/о человеке	ról/ről
11. Allativus	emberhez к человеку	hoz/hez/höz
12. Adessivus	embernél у человека	nál/nél
13. Ablativus	embertől от человека	tól/től

^{xx/} Во избежание новых терминов мы употребляем латинские названия, Порядок падежей несколько изменен по сравнению с [4].

Окончание:

14.	Terminativus	emberig до человека	ig
15.	Temporalis	öt ^{kor} в пять (часов)	kor
16.	Causalis-finalis	emberért за человека	ért
17.	Translativus- factivus	emberre человеком (стать)	vá/vé
18.	^{xxx/} Modalis-essivus I.	szomor ^{an} грустно	n/an/on/en/ön
19.	Modalis-essivus II. ^{xxx/}	egyidej ^{ül} eg одновременно	lag/leg
20.	Essivus-modalis	ember ^{ül} по-человечески	ul/ül
21.	Essivus-formalis	emberk ^{ént} как человек	ként
22.	Formalis	példak ^{épp} /példak ^{éppen} примером	képp/képpen
23.	Distributivus	fejenk ^{ént} (по...) с человека	nként/anként/onkén enként/önként
24.	Distributiv temporalis	nyaranta по летам	nta/anta/onta/ ente/önte
25.	Multiplicativus	öt ^{ször} в пять раз	szor/ször
26.	Locativus	Pécsett в городе Печ	t/ott/ett/ött
27.	Sociativus	kutyá ^{stul} вместе с собакой	stul/stül

^{xxx/} Большинство венгерских падежей может относиться и к существительному, и к прилагательному, но некоторые относятся только к прилагательному.

По огромному расхождению числа падежей в русском и венгерском языках понятно, (а по существу вытекает из чрезвычайной нагруженности функциями некоторых падежей русского языка), что отдельным падежам русского языка соответствует в венгерском языке целый ряд падежей в зависимости от характера и принадлежности корня, а также от синтаксической функции слова. Особенно много функций сочетают в себе родительный и творительный падежи русского языка. Рассмотрим, как разлагается в свой "спектр" творительный падеж русского языка без предлога "призмой" венгерского языка. Предварительно заметим, что с нашей точки зрения сущность языковых явлений выражается не только настоящим положением языка, поэтому для получения полной картины мы не отделяли застывшие выражения от живых. Кроме этого, в тех же целях включены в список и случаи, обусловленные просто управлением глагола. Единственным принципом нашего перечисления является установление соответствия русских падежных форм венгерским падежам или другим отыменным формам таким образом, чтобы разлагалась единая по форме категория творительного падежа по категориям венгерского языка.

1. a) Он пишет карандашом
Ceruzával ir val/vel
- б) Этот определитель получается
заменой коэффициентов
Ezt a determinánát az együtt-
hatók felcserélésével kapjuk
- в) весной, осенью
tavasszal, ősszel
2. a) зимой, летом
télen, nyáron n/on/en/ön
- б) путь, которым мы пройдем
az út, amelyen megyünk
- любой ценой
bármilyen áron

таким образом
ilyen módon

3. a) Последним они отличаются от них
Az utóbbiban különböznek azoktól ban/ben
- б) быстрыми темпами
gyors tempóban
- небольшими порциями
kis adagokban
4. пользуемся формулой
felhasználjuk a képletet t/at/ot/et/öt
5. Он был студентом
Diák volt ø
6. стало нетерпимым
tűrhetetlenné vált vá/vé
7. Такое множество называется пустым
Az ilyen halmazt üresnek nevezzük nak/nek
8. Он работал математиком
Matematikusként dolgozott ként
9. главным образом
főképp/főképpen képp/képpen
10. Он учился годами
Évekig tanult ig
11. поставить себе целью
célul tűzni maga elé ul/ül
12. лицом к лицу
szemtől szembe től/től

В) словообразовательный суффикс

ú/ü/ű/ű

В) последок

által

Г) аффикс и последок

n/on/en/ön át/keresztül

ra/re nézve

Полученная таким образом картина еще очень груба, однако, в общих чертах из нее уже вырисовываются - надо подчеркнуть: чисто формальным средством сопоставления соответствующих элементов - грамматические категории синтаксических функций, содержащиеся имманентно в творительном падеже русского языка.

Предложные выражения, хотя и они сохраняют в какой-то мере первоначальную функцию (или одну из первоначальных функций) данной падежной категории, гораздо менее интересны с этой точки зрения, поэтому не будем заниматься ими подробно.

Можно идти в противоположном направлении, рассматривая, какие элементы отношения соответствуют в русском языке венгерскому падежу *instrumentalis - comitativus*. Спектр в данном случае будет гораздо менее многообразным. Это и понятно, ведь отдельные падежи в венгерском языке, обладающем многими падежами, выполняют значительно меньше функций, чем падежи в русском языке.

1. Ceruzával írok
Я пишу карандашом

2. A barátommal sétálok
Я гуляю с другом

3. 30 %-kal növekedett a termelés
Производство повысилось на 30 процентов
4. Az x értéke 0-val egyenlő
Значение x равно нулю
5. x -szel jelöljük
Обозначается через x .

С особенной силой надо подчеркнуть важность 3-й, 4-й и 5-й функций с точки зрения математических текстов. Однако, можно резко ограничить эти функции лексически, пересчитывая списком случаи, при каких словах их употреблять. Например, функция 3 связана со следующими глаголами (или их отглагольными существительными):

növelni valamivel	увеличивать <u>на</u> что-нибудь
csökkenteni	уменьшать
szorozni	умножать
osztani	делить
egyszerűsíteni	сокращать
elmozdítani	сдвигать
felcserélni	заменять
különbözni	расходиться
bővíteni	расширять

4-я функция связана со следующими прилагательными или их производными:

egyenlő valamivel	равный <u>чему</u> -нибудь
egyenértékű	равносильный
ekvivalens	эквивалентный
párhuzamos	параллельный
szimmetrikus	симметричный
arányos	пропорциональный
ellentétes	противоположный
analóg	аналогичный
egynemű	однородный
izomorf	изоморфный

izometrikus
kongruens
kollineáris
komplanáris

изометричный
конгруэнтный
коллинеарный
компланарный

5-я функция связана с немногочисленными глаголами: глаголом обозначать, выражать, писать и их производными: обозначение, выражение и т.д.

Ниже мы попытаемся дать обзорную картину о соответствии элементов отношения русских и венгерских имен. Сбор материала проводился на основе прежде всего математических и физических текстов, а также с использованием данных венгерско-русского словаря Л.Хадровича и Л.Гальди [5]. Собранный материал мы проверили на основании выборки пробных текстов [6], и, исходя из этого, сделали небольшую статистику о частоте отдельных соответствий.

О принципах установления соответствия мы должны заметить следующие:

1. При выборе венгерских пар отдельных русских выражений мы не стремились дать красочные переводы, индивидуальные решения, наоборот, старались унифицировать соответствие, если это было возможно. Двойкий то подход к решению проблемы вытекает из принципиального различия характера человеческого и машинного перевода. Мыслящийся человек может во всех случаях взвешивать все обстоятельства, в том числе и смысловые связи, и на основании этого решать проблемы по-индивидуальному. Машина же, которая только имитирует мышление, дает всегда типовые в известном смысле решения, даже в случае, если ассортимент у нее очень богат.

Разницу двух подходов наглядно показывает следующий пример:

Словарь - помощник человека-переводчика - иллюстрирует одну из функций русского предлога "за" следующими примерами:

1) за недостатком "hiánya miatt", 2) за шумом ничего не слышно "a zajtól semmi sem hallható", 3) за отсутствием "távollétében" 4) за ненужностью "minthogy nincs rá szükség" "nem lévén rá szükség". Таким образом, словарь принимает пять разных способов перевода для четырех русских выражений. Однако, для машины различать эти возможности нет смысла, так как мы все равно получим удовлетворительный перевод во всех четырех примерах и в том случае, если поставим в соответствие предлогу "за" венгерский послелог "miatt": 1) hiánya miatt, 2) a zaj miatt, 3) távolléte miatt, 4) szükségtelensége miatt. Правда, перевод будет менее красочным, но соответствие элементов исследуемых языков становится более однозначным, отчетливым.

Нельзя не сказать, что практическое осуществление подобной унификации в некоторых случаях не очень простое дело.

Например, категория "о + винительный падеж" в русском языке функционально почти едина. А все-таки при переводе на венгерский язык встречается несколько вариантов ее выражения: ударяется о камень "nekiütözik a kőnek", "beleütözik a kőbe" или просто "kőbe ütődik" и т.д. Те же переводы правильны и для выражения "бьется головой о стену". Однако, другие примеры подсказывают другой перевод: Опирается руками о стол "Kezével az asztalnak támaszkodik". Ребята бьют мячом об пол "A gyerekek a padlóhoz ütögetik a labdát". Кажется, этот последний пример дает решение всех перечисленных и неперечисленных примеров, т.е. достаточно поставить в соответствие данному предлогу единственное венгерское окончание hoz/hez/höz (ablativus).

2) При выборе венгерских эквивалентов русских предлогов с некоторых точек зрения оказалось неподходящей проведенная выше система венгерских падежей, так что мы должны были произвести некоторые изменения в ней:

а) Кроме категории падежа "nominativus" употребляем и подкатегорию "nominativus-genitivus", кроме категории "dativus" и подкатегорию "dativus-genitivus". Причина этого заключается в следующем: В венгерском языке родительный падеж не имеет

особого окончания, он выражается или нулевым окончанием, или окончанием дательного падежа, но в обоих случаях формально распознаваема функция родительного падежа по лично-притяжательному окончанию а/е/ја/је "обладаемого". Например: книга мальчика "a gyerek könyve" или "a gyereknek a könyve". (Порядок обладателя и обладаемого не совпадает с порядком слов в русском языке).

б) Вместо падежа "distributivus-temporalis" (nte/nta/anta, onta/ente/önte) мы употребляем падеж "distributivus" (nként/anként/onként/enként/önként), которым может полностью заменяться предыдущее.

в) Вместо редко встречающегося падежа "locativus" (t/ott/ett/ött) употребляем падеж "superessivus" (n/on/en/ön) или падеж "inessivus" (ban/ben) соответственно (Pécsett → Pécsen, Győrött → Győrben).

г) Вместо редкого падежа "sociativus" (stul/stül) употребляем падеж "instrumentalis-comitativus" (val/vel).

3. Кроме падежей включены в список несколько типов прилагательных, а именно образованных от существительных с помощью суффиксов а) i б) s/ös/es/ös в) ü/ű/ju/jű, так как они выполняют аналогичную роль в венгерском языке, как существительные с предлогом в русском.

4. При составлении таблицы послелогов мы исходим из списка, составленного Денешем Сабо [7], но добавили к нему некоторые слова, которые более или менее уже можно считать послелогами.

5. Пока еще не проверено, почему и когда употребляется тот или другой венгерский падеж для выражения русских предложных или беспредложных оборотов. Пока эта работа не доведена, нам не хотелось отделить идиоматических оборотов и случаев глагольного управления от остальных, лексически не ограничиваемых случаев. Меньшая с точки зрения грамматики важность таких оборотов все-таки подчеркивается тем, что их знаки в сводных таблицах вставлены в скобках.

Венгерские падежи, соответствующие винительному
падежу русского языка

	Ø	в	за	на	о	по	под	про	сквозь	через
	A	B	C	D	E	F	G	H	I	J
1.1 Nominativus /ø/	+	+								
1.2 Nom.-gen. /ø/										
2. Accusativus /-t/	+	/+/	/+/	/+/		/+/				/+/
3.1 Dativus /-nak/		/+/	+	/+/			/+/			
3.2 Dat.-gen /-nak/										
4. Instr.-com. /-val/			/+/	+		/+/				+
5. Illativus /-ba/		+	/+/	+			/+/			
6. Inessivus /-ban/		+						/+/		/+/
7. Elativus /-ból/		/+/		/+/						
8. Sublativus /-ra/	/+/	+	/+/	+			/+/			/+/
9. Superessivus /-on/	+	+	/+/	+		+	/+/		/+/	/+/
10. Delativus /-ról/			/+/					+		
11. Allativus /-hoz/			+	/+/	+					
12. Adessivus /-nél/			/+/							
13. Ablativus /-től/										
14. Terminativus /-ig/	/+/					+				
15. Temporalis /-kor/		+								
16. Caus.fin. /-ért/			+	/+/		/+/				
17. Translat.-fact. /-vá/		+								
18. Mod.ess. I. /-an/										
19. Mod.ess. II. /-lag/										
20. Ess. mod. /-ul/		+		/+/			/+/	/+/		
21. Ess. form. /-ként/		+	+							
22. Formalis /-képpen/		/+/								
23. Distribu- tivus /-nként/	/+/	+		/+/		+				+
24. Multipli- cativus /-szor/		+								
25. Суфф.прил. I. /-i/		/+/	+	+						
26. " " II. /-os/		+		+			/+/			
27. " " III. /-ú/				/+/						

Примуры:

- 1.1 A Он работал целый день
Egész nap dolgozott
- 1.1 B Шар диаметром в 10 см
10 cm átmérőjű gömb
- 2 A Он читает книгу
Könyvet olvas
- 2 B звонить в колокол
a harangot megkondítani
- 2 C ступить за порог
átlépni a küszöböt
- 2 D опоздать на час
egy órát késni
- накрывать на стол
az asztalt megteríteni
- 2 J переступить через порог
átlépni a küszöböt
- 3.1 B идти в солдаты
katonának menni
- 3.1 C принять за единицу
egységnek venni
- 3.1 D Время работает на нас
Az idő nekünk dolgozik
- учиться на инженера
mérnöknek tanulni
- 3.1 G Приспособляют здание под школу
Iskolának átalakítják az épületet
- 4 C за пять дней до срока
5 nappal határidő előtt
- 4 D увеличить, уменьшить ... на 3
3-mal növelni, csökkenteni...

- 4 F по двое
kettesével
- 4 J выражать, обозначать, писать через x
x-szel fejezni ki, jelölni, írni
- 5 B в уравнение
az egyenletbe
- 5 C Он держится руками за перилу
A korlátba kapaszkodik.
- 5 D на фабрику
a gyárba
- 5 G попасть под дождь
esőbe kerülni
- 6 B в эпоху
korszakban
- 6 H читать про себя
magában olvasni
- 6 J через каждые два года
minden két évben
- 7 B во все горло
teli torokból
- 7 D на память
emlékezetből
- 8 A Он помнит это время
Emlékszik erre az időre
- 8 B в адрес
címre
- 8 C за три километра от города
3 km-re a várostól
- 8 D опираться на стол
az asztalra támaszkodni
- 8 G Здание отдано под оперные спектакли
Az épületet odaadták operai előadásokra

8	J	от завтрашнего дня, через две недели tegnaphoz két hétre
9	A	Он приходит каждую неделю Minden héten jön
9	B	в среду szerdán
9	C	выбросить за окно kidobni az ablakon
9	D	на третий день a harmadik napon
9	F	по другую сторону a másik oldalon
9	G	взять под руку kézen fogni
9	I	пробираться сквозь толпу áttörni a tömegben
9	J	Он прошел через все испытания Minden megpróbáltatáson keresztülment
10	C	выбросить за борт ledobni a fedélzetről
10	H	про всякий объект minden objektumról
11	C	приняться за работу munkához látni
11	D	дать своё согласие на что-н. beleegyezését adni valamihez
11	E	молекулы ударят о тело a testhez ütődnek a molekulák
14	A	Он работал долгое время Sokáig dolgozott
14	F	по январь januárig

- 15 B в час
egy órákor
- 16 C за мир
a békéért
- 16 D он купил книг на 30 рублей
30 rubelért könyvet vásárolt
- 16 F идти по воду
vizért menni
- 17 B превратиться во что-нибудь
átváltozni valamivé
- 20 B в заключение
befejezésül
- 20 G использовать под базы
bázisul használni
- 20 H про запас
tartalékul
- 21 B в ответ
feleletként
- 21 C принять за правило
szabályként elfogadni
- 22 B не в обиду будь сказано
nem sértésképpen mondva
- 23 A Журнал выходит каждую неделю
A folyóirat hetenként jelenik meg
- 23 B в неделю раз
hetenként egyszer
- в час
óránként
- 23 D по сколько на каждого?
fejenként hányat?
- 23 J через два часа
2 óránként

- 24 B в три раза больше
ötször több
в пятый раз
ötödször
- 25 B иск в суд
bírószági kereset
- 25 C заработок за год
egy évi bér
- 25 D план на этот год
ezévi terv
отпуск на месяц
egy hónapi szabadság
- 26 B бомба весом в тонну
egy tonnás bomba
- 26 D обед на двадцать персон
20 személyes ebéd
- 26 G бутылка под молоко
tejés üveg
- 27 D на бытовые темы
hétköznapi témáju

Венгерские падежи, соответствующие родительному
падежу русского языка

	∅	в виде	в каче- стве	внутри	внутри	вокруг	для	до	из	из-за	из-под	на счет	от	против	с	среди	у
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1.1 Nominativus /∅/	+															/+/	
1.2 Nom.-gen. /∅/	+								+				+				
2. Accusativus /-t/	+																
3.1 Dativus /-nak/																	+
3.2 Dat.-gen. /-nak/	+								+				+				+
4. Instr.-com. /-val/								/+/							+		
5. Illativus /-ba/					/+/								/+/				
6. Inessivus /-ban/				+									/+/				
7. Elativus /-ból/	+						/+/		+	/+/			/+/		+		
8. Sublativus /-ra/						/+/	+	+									
9. Superessivus /-on/	+															/+/,	
10. Delativus /-ról/						/+/			/+/	/+/		+	/+/		+		
11. Allativus /-hoz/							+						/+/				/+
12. Adessivus /-nál/	+						/+/							/+/			+
13. Ablativus /-től/	/+/									/+/			+		+		+
14. Terminativus /-ig/								+									
15. Temporalis /-kor/																	
16. Caus.fin. /-ért/							/+/										
17. Translat.-fact./-vá/																	
18. Mod.ess. I. /-an/							/+/										
19. Mod.ess.II. /-lag/																	
20. Ess.mod. /-ul/																	
21. Ess.form. /-ként/		+	+														
22. Formalis /-képpen/																	
23. Distributivus /-nként/																	
24. Multiplicativus /-szor/																	
25. Adj.suff. I. /-i/	+												+				
26. Adj.suff. II. /-os/							+		+		/+/	/+/					
27. Adj.suff. III./-u/																	

Примеры:

- 1.1 A Петра нет дома
Nincs itthon Péter
- 1.1 P Он проснулся среди ночи
Éjszaka felébredt
- 1.2 A теория вещественных чисел
a valós számok elmélete
- 1.2 I один из студентов
a hallgatók egyike
- 1.2 M функция от x
 x függvénye
- 1.2 Q ножка у стула
a szék lába
- 3.1 Q у Петра есть книга
Péternek van könyve
- 2 A Я не получил газеты
Nem kaptam ujságot
- 3.2 A Случай равенства нулю определителей
A determinánsok nullával való egyenlőségének esete
- 3.2 I один из студентов
egyike a hallgatóknak
- 3.2 M лежат по одну сторону от линии
a vonalnak az egyik oldalára esnek
- 3.2 Q Дверь у шкафа заперта
A szekrénynek az ajtaja be van zárva
- 4 H У меня до тебя дело
Dolgom van veled
- 4 O начать с большой буквы
nagy betűvel kezdeni
- встать с левой ноги
bal lábbal kelni
- с позволения кого-нибудь
valakinek az engedélyével

5	E	внутри атома az atom belsejébe
5	M	он умер от раны belehalt a sebbe
6	D	внутри главного здания a főépületben
6	M	от имени nevében
7	A	Я выпил воды ittam a vizből
7	G	для этой цели e célból
7	I	выйти из дому kijönni a házból
7	J	выхватить кинжал из-за пояса kirántani a tőrten az övéből
7	M	от сердца szívből
7	O	прийти с завода az üzemből jönni
8	F	огибать обруч вокруг бочки ráhuzni az abroncsot a hordóra
8	G	для х мы получим х-re azt kapjuk, hogy...
8	H	увеличить, уменьшить, округлять, дополнить, снизиться, разгонять до чего-нибудь növelni, csökkenteni, kerekíteni, kiegészíteni, csökkenni, felgyorsítani valamire
		частицы сближаются до расстояния Т a részecskék Т távolságra közelednek

- 9 A первого мая
május elsején
- 9 P среди ясного неба появились тяжелые темные тучи
A világos égen sűrű, sötét felhők jelentek meg
- 10 F вокруг Чехова
Csehovról
- 10 I из года в год
évről évre
- 10 L Вы как думаете насчет этого?
Ön mit gondol erről?
- 10 M от времени до времени
időről időre
- от слова до слова
szóról szóra
- 10 O с одной стороны на другую
egyik oldalról a másikra
- 11 G для этого необходимо
ehhez szükséges
- 11 M близко от города
közel a városhoz
- 11 Q поставить стол у окна
az ablakhoz tenni az asztalt
- 12 A сильнее Петра
erősebb Péternél
- 12 G Для него необычно приходить так поздно
Nála szokatlan, hogy ilyen későn jöjjön
- 12 N Мое пальто против твоего гораздо лучше
Az én kabátom sokkal jobb a tiednél
- 12 Q у двери
az ajtónál
- 13 A Запись, лишенная математического содержания
Matematikai tartalmától megfosztott felírás

13	J	встать из-за стола felkelni az asztaltól
13	M	от ияти часов öt órától
13	O	с трех до ияти háromtól ötig
13	Q	спросить у кого-нибудь megkérdezni valakitől
14	H	с трех до пяти háromtól ötig до известной степени bizonyos fokig
16	G	для того, чтобы azért, hogy
18	H	тощий до смешного nevetségesen sovány
21	B	в качестве первого примера első példaként
25	A	состояние покоя nyugalmi állapot
25	M	постановление от первого февраля február elsejei határozat
26	G	ящик для чая teásdoboz
26	I	обед из двух блюд kétfogásos ebéd
26	J	коробка из-под конфет cukrosdoboz

Венгерские падежи, соответствующие дательному и
творительному падежам русского языка

	Дательный падеж			Творительный падеж						
	Ø	к	по	Ø	за	между	с	над	перед	по
	A	B	C	A	B	C	D	E	F	G
1.1 Nominativus /ø/	/+/			+			/+/			
1.2 Nom.-gen. /ø/	+							/+/		/+
2 Accusativus /-t/	/+/		/+/	/+/	/+/		/+/			
3.1 Dativus /-nak/	+	/+/		+						
3.2 Dat.-gen. /-nak/	/+/									
4. Instr.-com. /-val/	+		/+/	+	/+/		+	+		/+
5. Illativus /-ba/			/+/							
6. Inessivus /-ban/			+	/+/						/+
7. Elativus /-ból/			/+/							
8. Sublativus /-ra/	+	+	/+/	/+/	/+/					
9. Superessivus /-on/			+	/+/				+		
10. Delativus /-ról/			/+/							
11. Allativus /-hoz/	/+/	+				/+/				
12. Adessivus /-nál/					/+/					+
13. Ablativus /-től/	/+/		/+/	/+/					/+/	
14. Terminativus /-ig/			/+/	/+/						
15. Temporalis /-kor/										
16. Caus.fin. /-ért/					+					
17. Translat.-fact. /-vá/				+						
18. Mod.ess. I. /-an/			/+/							
19. Mod.ess.II. /-lag/			/+/							
20. Ess.mod. /-ul/		/+/	+	+						
21. Ess.form. /-ként/										
22. Formalis /-képpen/										
23. Distributivus /-nként/			+							
24. Multiplicativus /-szor/				+						
25. Adj.suff. I. /-i/			+							+
26. Adj.suff. II. /-os/			+				+			
27. Adj.suff.III. /-ú/			+	+			+			

Примеры:

Дательный падеж

- 1.1 A Ивану хотелось бы уйти
Iván szeretett volna elmenni
- 1.2 A памятник Пушкину
Puskin emlékműve
- 2 A Эти многогранники удовлетворяют требованиям
Ezek a poliéderek kielégítik a feltételeket
- 2 C погладить по головке
megsimogatni a fejét
- 3.1 A Она дала ему книгу
Adott neki egy könyvet
- 3.1 B К чему он сказал?
Minek mondta?
- 3.2 A Это - памятник Пушкину
Ez Puskinnak az emlékműve
- 4 A равно нулю
egyenlő nullával
- 4 C по пяти
ötösével
- 5 C проваживать ребят по театрам
kisérgetni a gyerekeket a színházakba
- 6 C Работа велась по трем направлениям
Három irányban folyt a munka
- 7 C по этой причине
ebből az okból
сдать экзамен по математике
vizsgázni matematikából
- 8 A перпендикулярный прямой
merőleges az egyenesre

- 8 B перейти к следующему вопросу
áttérni a következő kérdésre
- 8 C развесить картины по стенам
felaggatni a képeket a falra
- 9 C говорить по телефону
telefonon beszélni
- 10 C я сказал все по данному вопросу
Mindent elmondtam az adott kérdésről
- 11 A он принадлежит множеству
a halmazhoz tartozik
- 11 B x относится к y
x aránylik az y-hoz
- 13 A чуждый ему
idegen tőle
- 13 C По мне как хочешь
Tőlem - ahogy akarod
- 14 C Я не видел ее по целым неделям
Hetekig nem láttam
- 19 C они отличаются лишь по виду
csak formailag különböznek
- 20 B к примеру
például
- 20 C по-русски
oroszul
- 23 C по временам
időnként
- по частям
részenként
- 25 C исследования по геометрии
geometriai vizsgálatok
- 26 C мальчик по двенадцатому году
tizenkétéves gyerek

- 27 C одинаковые по массе
egyenlő tömegűek
- Они содержат по одинаковому числу атомов
egyenlő számú atomot tartalmaznak

Творительный падеж

- 1.1 A Она была студенткой
Diák volt
- 1.1 D с каждым днем
minden nap
- 1.2 E мошенник над мошенниками
csalók csalója
- 1.2 поле под культурными растениями
kulturnövények földje
- 2 A Мы пользуемся теоремой Коши
A Cauchy-tételt használjuk.
- 2 B следить за его мыслью
követni a gondolatát
- 2 D с новым годом
boldog új évet
- 3.1 A считаем его неизменным
tekintsük változatlanak
- 4 A писать карандашом
ceruzával írni
- 4 B за немногими исключениями
néhány kivétellel
- 4 D с неизменными условиями
változatlan feltételekkel
- 4 E действия над вещественными числами
műveletek valós számokkal
- 4 G рыба под соусом
hal szósszal

6	A	отличаться чем-нибудь különböző valamiben сомкнутыми рядами zárt sorokban
6	G	пересекать под равными углами egyenlő szögben metszeni
8	A	родом он венгерец születésére magyar
8	B	наблюдать за детьми felügyelni a gyerekekre
9	A	зимой, летом télen, nyáron
9	E	работать над книгой dolgozni a könyvön
11	C	сходство языков между собой a nyelvek hasonlósága egymáshoz
12	B	за столом az asztalnál
12	G	нет микрометра под рукой nincs kéznél mikrométer
13	F	извиниться перед кем-нибудь bocsánatot kérni valakitől
14	A	Он учился годами Évekig tanult
16	B	послать за доктором az orvosért küldeni
17	A	стало необходимым elkerülhetetlenné vált
20	A	Мы оставили неизменным значение параметра Változatlanul hagytuk a paraméter értékét
24	A	пятью шесть öttször hat

- 25 G битва под Москвой
a moszkvai csata
- 26 D уравнение с тремя неизвестными
háromismeretlenes egyenlet
- 27 A шар диаметром в 10 см
10 cm átmérőjű gömb
- 27 D элемент с порядковым номером 98
a 98-as rendszámú elem

Венгерские падежи, соответствующие предложному
падежу русского языка

- 1.1 Nominativus /ø/
 1.2 Nom.-gen. /ø/
 2. Accusativus /-t/
 3.1 Dativus /-nak/
 3.2 Dat.-gen. /-nak/
 4. Instr.-com. /-val/
 5. Illativus /-ba/
 6. Inessivus /-ban/
 7. Elativus /-ból/
 8. Sublativus /-ra/
 9. Superessivus /-on/
 10. Delativus /-ról/
 11. Allativus /-hoz/
 12. Adessivus /-nál/
 13. Ablativus /-től/
 14. Terminativus /-ig/
 15. Temporalis /-kor/
 16. Caus.-fin. /-ért/
 17. Translat.fact. /-vá/
 18. Mod.ess. I. /-an/
 19. Mod.ess. II. /-lag/
 20. Ess.mod. /-val/
 21. Ess.form. /-ként/
 22. Formalis /-képpen/
 23. Distributivus /-nként/
 24. Multiplicativus /-szor/ - - - -
 25. Adj.suff. I. /-i/
 26. Adj.suff. II. /-os/
 27. Adj.suff. III. /-ű/

Ø	в	на	о	при
A	B	C	D	E
	/+/ +		+	+
	/+/ +			
	/+/ +		+	+
	/+/ +	/+/ +		+
	+	+		/+/ +
	/+/ +			
	/+/ +	/+/ +	/+/ +	
	+	+		
	/+/ +		+	
		/+/ +		
				+
	+			+
			/+/ +	
	+			
	+			
	+	+		+
	+	+		
	+	+	+	

Примеры:

- 1.2 D понятие об определителе
a determináns fogalma
вопрос о линейности
a linearitás kérdése
- 1.2 E коэффициент при x
x együtthatója
- 2 B признаться в ошибках
beismerni a hibáit
отказывать в помощи, в просьбе
megtagadni a segítséget, kérést
- 4 B обвинять кого-нибудь в чем-нибудь
vádolni valakit valamivel
- 4 C не на словах, а на деле
nem szóval, hanem tettel
- 4 E при таких условиях
ilyen feltételekkel
- 6 B в уравнении
az egyenletben
- 6 C на заводе
az üzemben
на практике
a gyakorlatban
- 6 E при жизни Менделеева
Mendelejev életében
- 7 B в целях упрощения
egyszerűsítés céljából
- 8 B в трех километрах от города
3 km-re a várostól
- 8 C на двадцати км от Москвы
20 km-re Moszkvától

- 8 D Я не думал об этом
Nem gondoltam erre
- 9 B в квадрате
a négyzeten
- 9 C на разных языках
különböző nyelveken
- 10 B убедиться в чем-нибудь
meggyőződni valamiről
- 10 D замечания об основных теоремах
megjegyzések a fő tételekről
- 11 C настаивать на убеждениях
ragaszkodni a meggyőződéséhez
- 12 E при положительных значениях x
 x pozitív értékeinél
- 15 B в половине четвертого
félnégykor
- 15 E при изменении параметра
a paraméter megváltoztatásakor
- 16 D О чем ты плачешь?
Miért sírsz?
- 21 B в результате
eredményként
- 24 B во-первых
először
- 25 B объяснение в любви
szerelmi vallomás
- 25 C Ростов на Дону
a doni Rosztov
- 25 E ясли при заводе
üzemi bölcsőde
- 26 B дом в пяти этажах
ötemeletes ház

- 26 C раствор на дистиллированной воде
desztillált vizes oldat
- 27 B боксер в легком весе
könnyű súlyú boxoló
- 27 C книги на русском языке
orosz nyelvű könyvek
- 27 палка о двух концах
két végű bot

Выражение венгерских послелогов с помощью предлогов
русского языка

	Послелог	Винительный п.	Родительный п.
1.	alá	под, сквозь	
2.	alatt	за, в	
3.	alól		из-под
4.	fölé		сверх
5.	fölött		сверх
6.	fölül		
7.	elé	над, за, в	
8.	előtt	над	до
9.	elől		от, с, из
10.	mögé	за	
11.	mögött	за	позади
12.	mögül		из-за
13.	mellé	за	около, возле, у
14.	mellett	на, за, под	около, возле, у
15.	mellől		от, с
16.	közé	в	между, среди, к числу
17.	között	в, под, сквозь	между, среди, в числе
18.	közül		из, из-под
19.	köré		вокруг
20.	körül		вокруг, около
21.	felé	под, на	около
22.	felől		со стороны

	Послелог	Дательный падеж	Творительный падеж	Предложный падеж
1.	alá			
2.	alatt		под	на, при
3.	alól			
4.	fölé		над	
5.	fölött		над	
6.	fölül			
7.	elé	навстречу	перед	
8.	előtt	назад	перед	на, при
9.	elől		перед	
10.	mögé			
11.	mögött	вслед	за	
12.	mögül			
13.	mellé			
14.	mellett		наряду с, за	на, при
15.	mellől			
16.	közé	к	между	
17.	között		между	в
18.	közül			
19.	köré			
20.	körül			
21.	felé	к		
22.	felől	по		
23.	után	вслед, по	за, вслед за	по
24.	iránt	к, по на- правлению к	перед	
25.	ellen		над, под	

	Послелог	Винительный п.	Родительный падеж
23.	után	за, под	после, от, по истечении
24.	iránt		
25.	ellen	на, под	против, напротив, от
26.	mulva	через	без, по истечении
27.	óta		с
28.	közben	под	во время, среди, в течение, с течением
29.	helyett	за	вместо, взамен
30.	szerint		
31.	által	про	
32.	nélkül		без
33.	miatt	за	из-за, от, с, ради
34.	lévén		
35.	esetén		в случае
36.	számára /részére/	на	для
37.	ellenében	за, под, на	против, взамен
38.	ellenére	несмотря на	вопреки, против, сверх, вне
39.	következtében		вследствие, из-за
40.	révén	за, через	за счет, путем, посредством, при посредстве, через посредство, с помощью, при помощи
41.	folyamán		среди, в течение, с течением
42.	nyomán	за	
43.	terén		в области
44.	céljából		для, с целью
45.	értelmében		в силу
46.	érdekében		в пользу
47.	irányában		по линии

	Послелог	Дательный падеж	Творительный падеж	Предложный падеж
26.	mulva			в... /отсюда/
27.	óta			
28.	közben		за, между	при, на
29.	helyett			
30.	szerint	по, согласно		в, на
31.	által			
32.	nélkül			
33.	miatt	по	за	о
34.	lévén		будучи	
35.	esetén			при
36.	számára/részére/			
37.	ellenében			
38.	ellenére			при
39.	következtében	благодаря, по		
40.	révén			при
41.	folyamán			
42.	nyomán	по	вслед за, за	
43.	terén			
44.	céljából			
45.	értelmében	по, согласно		
46.	érdekében			
47.	irányában	по, по на- правлению к		
48.	végére			
49.	alkalmával			
50.	mértékében			

	По́сле́лог	Винительный п.	Родительный падеж
48.	végére		к концу, в конце
49.	alkalmával		по поводу
50.	mértékében		по мере
51.	-szerte		
52.	utján		путём
53.	módra		
54.	segítségével		при помощи, с помощью
55.	módon		
56.	-on alul		ниже, до
57.	-on felül	за	свыше, сверх, выше, вне
58.	-on belül	в	внутри
59.	-on kívül		без, вне, кроме, сверх
60.	-on át	через, сквозь	в течение
61.	-on innen		по эту сторону,
62.	-on túl	за	по ту сторону, свыше, сверх, вне, дальше
63.	-ra nézve		относительно, для
64.	-ra vonatkozólag		для, насчет
65.	-ra alkalmazásával		
66.	-tól távol		далеко от, вдали от, вдалеке от
67.	-tól fogva		с, начиная с
68.	-tól eltekintve		кроме, без
69.	-nál fogva		из-за, от, ввиду
70.	-val együtt		
71.	-val szemben		напротив, против
72.	-val kapcsolatban		по поводу, относительно

	Послелог	Дательный падеж	Творительный падеж	Предложный падеж
51.	-szerte	по		
52.	utján			
53.	módra	по		
54.	segítségével	по		
55.	módon	по-	...образом, ...способом	
56.	-on alul			
57.	-on felül			
58.	-on belül			
59.	-on kívül			
60.	-on át	по		
61.	-on innen		под, за	
62.	-on túl		за	
63.	-ra nézve	по	в связи с	
64.	-ra vonatkozólag			
65.	-ra alkalmazásával	применительно к		
66.	-től távol			
67.	-től fogva			
68.	-től eltekintve			
69.	-nál fogva	по	под	
70.	-val együtt		вместе с, наряду с	
71.	-val szemben	навстречу, к		
72.	-val kapcsolatban	по отношению к	в связи с	
73.	-hoz képest	по отношению к	по сравнению с	

	Послелог	Винительный п.	Родительный падеж
73.	-hoz képest		для, против
74.	-hoz közel	под	вблизи, близ, около
75.	-nak javára		для пользы
76.	-nak körében		среди
77.	-nak megfelelően		

Послелог	Дательный падеж	Творительный падеж	Предложный падеж
74. -hoz közel		перед	
75. -nak javára			
76. -nak körében			
77. -nak megfelelően	согласно, по		

Соответствие венгерских падежей /суффиксов, после-
логов/ и беспредложных оборотов русского языка

	Название	Окончания	Падежи русского языка				
	венгерских падежей		И	В	Р	Д	Т
1.1	Nominativus	∅	+	/+ /	+	/+ /	/+ /
1.2	Nom.-gen.	∅			+		
2.	Accusativus	-t	/+ /	+	+	/+ /	/+ /
3.1	Dativus	-nak				+	/+ /
3.2	Dat.-gen.	-nak			+		
4.	Instrumentalis- comitativus	-val				+	
5.	Illativus	-ba					
6.	Inessivus	-ban					/+ /
7.	Elativus	-ból			+		
8.	Sublativus	-ra		/+ /		/+ /	/+ /
9.	Superessivus	-on		/+ /	/+ /		/+ /
10.	Delativus	-ról					
11.	Allativus	-hoz				/+ /	
12.	Adessivus	-nál			/+ /		
13.	Ablativus	-től			/+ /	/+ /	/+ /
14.	Terminativus	-ig		/+ /			/+ /
15.	Temporalis	-kor					
16.	Causalis-finalis	-ért					
17.	Translativus-factivus	-vá					+
18.	Modalis essivus I.	-an					/+ /
19.	Modalis essivus II.	-lag					

	Название	Окончания	И	В	Р	Д	Т
20.	Essivus modalis	-ul					+
21.	Essivus formalis	-ként					+
22.	Formalis	-képp/en/					
23.	Distributivus	-nként		/ + /			
24.	Multiplicativus	-szor					+
25.	Adj.suff. I.	-i			+		
26.	" " II.	-os			/ + /		
27.	" " III.	-ú			+		+
28.	Postpositio	által					+
29.	" "	alatt					/ + /
30.	" "	elé				/ + /	
31.	" "	-ra nézve					+
32.	" "	-on át/keresztül		+			+
33.	" "	-hoz képest				+	
34.	" "	segítségével, útján					+

Проверка материала на основании пробных текстов
математического характера

1. Встречаемость предложных оборотов:

Номер текста	Число букв	Число слов	Число предлогов	Процентное отношение предлогов
1.	18 000	2371	205	8,6
2.	10 400	1443	84	5,8
3.	9 200	1264	139	11,0
4.	7 400	949	83	8,8
5.	5 000	672	73	10,9
	<hr/> 50 000	<hr/> 6696 слов	<hr/> 584 предлога	<hr/> 8,7 %

2. Перевод предложных оборотов

Число предложных оборотов	При помощи падежей венгерского языка	При помощи послелогов
584	507 (86,8 %)	77 (13 2 %)

Распределение предложных оборотов по
падежам венгерского языка

	Венгерские падежи, соответствующие русским падежам	Винительный п.					Родительный п.							
		в	через	на	за	про	из	для	от	до	с	в ка- честве	у	внутри
1.1	Nominativus \emptyset													
1.2	Nom.-gen. \emptyset						25		1				1	
2.	Accusativus -t								2		3			
3.1	Dativus -nak				2									
3.2	Dat.-gen. -nak													
4.	Instr.-com. -val		12	7							1			
5.	Illativus -ba	11												
6.	Inessivus -ban	10												1
7.	Elativus -ból						29		1		1			
8.	Sublativus -ra	1		4				38		2				
9.	Superessivus-on													
10.	Delativus -ról					1								
11.	Allativus -hoz							10						
12.	Adessivus -nál							1						
13.	Ablativus -től								29					
14.	Terminativus -ig									4				
15.	Temporālis -kor													
16.	Caus.-fin. -ért							1						
17.	Transl.fact.-vá	2												
18.	Mod.-ess.I. -an								1					
19.	Mod.-ess.II.-lag													
20.	Ess.-mod. -ul	1												
21.	Ess.-form. -ként											3		
22.	Formalis -képp/en/													
23.	Distributivus -nként													
24.	Multiplicativus -szor													
25.	Adj.suff.I. -i													
26.	Adj.suff.II. -os													
27.	Adj.suff.III. -ú													

распределение русских
падежей с предлогом,
соответствующих вен-
герским падежам

25 12 11 2 1
51
(10,1%)

54 50 33 6 5 3 1 1
153
(30,0%)

	Венгерские падежи, соответствующие русским падежам	Дат. п.		Твор.п.			Предл.п.				распреде- ние вен- герских падежей	В % -ах
		к	по	с	за	под	в	при	о	на		
1.1	Nominativus \emptyset											
1.2	Nom.-gen. \emptyset							5			32	6,3
2.	Accusativus -t										5	1,0
3.1	Dativus -nak										2	0,4
3.2	Dat.-gen. -nak						1	1	2		4	0,8
4.	Instr.-com. -val		2	29	1						52	10,3
5.	Illativus -ba										11	2,2
6.	Inessivus -ban		2			1	113			3	130	25,7
7.	Elativus -ból		1								31	6,1
8.	Sublativus -ra	10	1								56	11,0
9.	Superessivus -on		1				4			30	35	6,8
10.	Delativus -ról								31		32	6,3
11.	Allativus -hoz	16									26	5,1
12.	Adessivus -nál							25			26	5,1
13.	Ablativus -tól										29	5,7
14.	Terminativus -ig										4	0,8
15.	Temporalis -kor							11			11	2,2
16.	Caus.fin. -ért										1	0,2
17.	Transl.fact. -vá										2	0,4
18.	Mod.ess.I. -an										1	0,2
19.	Mod.ess.II. -lag		1								1	0,2
20.	Ess.mod. -úl										1	0,2
21.	Ess.form. -ként										3	0,6
22.	Formalis -képp(en)										-	
23.	Distributivus -nként										-	
24.	Multiplicativus -szor						4				4	0,8
25.	Adj.suff.I. -i										-	
26.	Adj.suff.II. -os			3							3	0,6
27.	Adj.suff.III. -ű			5							5	1,0

26 8 37 1 1 122 42 33 33 507 100%
 34 39 230
 (6,7%) (7,7%) (45,5%)

ЛИТЕРАТУРА

- [1] МЕЛЬЧУК, И.А., О машинном переводе с венгерского языка на русский, в Проблемы Кибернетики I, р. 222, 1958
- [2] BALASSA, J., Книга венгерского языка^{*}/Будапешт, 1943/; русское издание: Венгерский язык/Москва, 1951/, р.254.
- [3] ANTAL, L., Система венгерских падежей^{*}/Будапешт, 1961/; /17 падежей/ Майтинская, К.Е., Венгерский язык /Москва, 1955/; /22 падежа/
- SZABÓ, D., Современный венгерский язык, конспект для филологического факультета университета^{*} /Будапешт, 1954/; /23 падежа/
- TOMPA, J. /ред./, Система современного венгерского языка, том I^{*}/Будапешт, 1961/; /28 падежей/
- [4] Система современного венгерского языка /см. выше/
- [5] HADROVICS, L. - GÁLDI, L., Русско-венгерский словарь, 2-е изд. /Будапешт, 1959/
- [6] СМЕРНОВ, В.И., Курс высшей математики, т. III/1 /Москва, 1952/; рр. 36-47.
- ХИНЧИН, А.Я., Краткий курс математического анализа /Москва, 1953/, рр. 76-80.
- ГИЛЬБЕРТ, А., Основания геометрии /Москва, 1948/, рр. 85-91.
- БОНГАРД, М.М., О понятии полезной информации, Проблемы Кибернетики 9, рр. 71-74, 1963
- ГНЕДЕНКО, В.В., Курс теории вероятностей, /Москва, 1950/, рр. 71-73.

^{*} На венгерском языке

- [7] SZABÓ, D., Современный венгерский язык /см. выше/, р.157;
таблица дополнена на основании Топра, J., Система
современного венгерского языка /см. выше/

SOME PROBLEMS OF SYNTHESIS OF THE HUNGARIAN NOUN FORMS

by J. Klauszer

In this brief article a survey is given of the problems concerning the synthesis of the Hungarian noun forms and also the method how to tackle it. It is to be born in mind that a Hungarian noun may have 714 paradigmatical (declension-) forms. (Cf. Antal A magyar esetrendszer, Budapest 1961. - p.50.) Apart from the nouns ending in vowels we have completed the algorithms of nouns with 11 terminal consonants, and on the basis of these investigations we have drawn some useful conclusions.

A few words must, however, be said about the difficulties encountered in our work. Since we do not have an a tergo dictionary at our disposal, and it is the ending that has to be taken as the basis, we have attempted to overcome this difficulty in the following way.

On the basis of a frequency dictionary given in the book "Szóstatisztika egymillió szótagot felölelő műveltségsszövegek alapján" by Zoltán Nemes, Ph.D. (EMGYK, Vol. 190, 1941) a micro a tergo dictionary containing approximately 700 nouns has been compiled taking into consideration the last 4 or 5 letters of each item. This was our material to start with, but in itself it was found far too meagre. For this reason we supplemented this initial stock of words partly with our own examples, partly with the help of a rhyming dictionary compiled by Kristóf Simai in 1809. In this way we succeeded in collecting a fair number of words which seemed to be satisfactory.

The other main obstacle is the lack of a detailed structural description of the Hungarian language. To make up for this Hungarian textbooks for foreigners have been used with profit. Many interesting data are provided by János Lotz's book "Das ungarische Sprachsystem" (Stockholm, 1939). Special mention must

also be made of K.E. Maytinskaya's book "Vengersky yazik" (Moscow, 1959) She discusses the characteristics of the Hungarian language in great detail, in connection with nouns and states several rules not yet mentioned in any other book so far. These grammar books, however, even if they are of great help cannot fulfill the need for a structural (descriptive) grammar.

Brief mention should also be made of our work so far. It seemed necessary to draw up a system of declensions for the Hungarian nouns too. We tried to find out those "crucial" cases which could serve as a basis for systematization. The following three cases proved to be most suitable for this purpose: Accusative Singular, Nominative Plural, and 3d person Singular of the Possessive Case. It is these cases where the characteristic changes in noun declension take place (changes in stem, problem of linking vowels, etc.). So including the palatal-velar variants we got 21 declension types as a result. We found a palatal pair for each of the velar types but there we came upon a palatal type without velar pair, thus palatals proved to be "larger", and velars had to be telescoped into them.

Since we use only palatal declension types at the beginning terminal velar vowels are to be transcribed into palatal ones. Therefore we have to get as a result forms such as: házeinek, kalapöknek, alméknek, etc. To eliminate such forms at the end we apply - in compliance with the rule of vowel harmony - an algorithm for replacing the palatal endings with velar ones after roots of type "velar".

The types of declension are the following:

Cons.

I. or type E

- | | |
|----------|---------|
| 1) t ek | (t ak/ |
| 2) t ök | (t ok) |
| 3) et ek | (at ak) |
| 4) öt ök | (ot ok) |

II. or type JE

- | |
|------------------|
| 1) t ek (--) |
| 2) t ök (t ok) |
| 3) et ek (at ak) |
| 4) öt ök (ot ok) |

Vow.

1) ǫ, ū, ŭ (ó, u, ú)	t je k (t ja k)
2) i (i)	t je k (t ja k)
3) e (a)	t je k (t ja k)

How can we decide whether a noun belongs to this or that of the declension types? The following considerations give some help:

- What does the word terminate in (u_1)
 - 1) in a vowel or in a consonant?
 - 2) in which of the vowels or consonants?
- What is there before the last letter of the word? (u_2).
- Is the word monosyllabic or polysyllabic?
- Is the word of palatal or velar harmony?
- If palatal, is it labial or illabial?

As it has already been mentioned above the algorithm of nouns which have a vocalic ending and of nouns with one of 11 terminal consonants has been completed.

Before drawing up the system of declensions a fair amount of data was collected for the individual letter-endings and suffixes. Having done so we could proceed only in one direction; that is the grouping of nouns with a vocalic or consonantal ending could have been done only in compliance with one certain principle. For instance: it is rather easy to show in the Possessive Case 3 Sg the desinences where -j- appears and those too where the possessive affix is only a/e. Moreover, there are some rules for the linking vowels of Nominative Plural and Accusative Singular. The rules, however, overlap when applied to the individual letters, i.e. they may apply to both.

In the determination of declension types - when we were considering only the three "crucial" cases - this made no trouble. But for setting up or algorithm new relationships must be pointed out, - which have not been spoken about so far - because cha-

racterisation or grouping of nouns on the basis of their endings is not possible in any other way.

There have already been here some exceptions to the rules pointed out above. They were treated separately as part of an exception chart (R in our notation). The full R-chart can only be drawn up when we have algorithms for every letter-ending. We have elaborated them only in parts but we consider it as complete and take them into consideration when working with algorithms.

When speaking about algorithms we have to treat nouns according to new principles. It follows that with new rules we have also to account for new exceptions. We have already met such "new" exceptions. They were temporarily marked at the algorithms and will be included in the R-chart containing all the data. If the whole a tergo vocabulary will be made up the amount of items on R will obviously increase.

Let us now in detail see the algorithms that have been completed so far and examine which of the above mentioned principles had to be taken into consideration. It is common in all of them that first we ask for R excluding the exceptions right at the beginning. It is more to the purpose to finish with the factors of little importance first and then go over to those which characterize the individual letter-endings. Double letters - such as sz, zs, cs, etc. - are considered as one.

S: this is perhaps one of the most complex of algorithms. Nouns ending in -s belong without exception to the first declension, and it follows that in the Possessive Case we have a or e in 3d person Singular. They can be declined according to all the four sub-types. To decide, on the other hand, that a noun with an -s in the end belongs to this or that of the sub-types we have to examine:

- What is u_2 ? In case it is a consonant, is it r?
- Which vowel is u_3 ?
- Is the word palatal or velar?

According to this:

I/1 is the type of declension if u_2 is e or é or i,
and the word is palatal.

I/2 if u_2 is either a or ă or o or u, or ö, or ü,
and if u_2 is i and the word is velar.

Exceptions: vas, has, magas.

I/3 if u_2 is r and u_3 is á or e or i

I/4 if u_2 is r and u_3 is o or ö

if u_2 is not r, but any other consonant

Sz: these nouns also belong to the first declension. Its algorithm is simpler because here we have only to examine what u_2 is and whether the word is of palatal or velar harmony.

Thus:

I/1 if the word is palatal and u_2 is a vowel

I/2 if the word is velar and u_2 is a vowel

I/3 if the word is palatal and u_2 is a consonant

I/4 if the word is velar and u_2 is a consonant

Z: these nouns also belong to the first declension. To be examined:

- What u_2 is

- Whether the word is monosyllabic or polysyllabic

I/1 if u_2 is e or é (exception: méz)

I/2 if the word is polysyllabic and u_2 is á or a

I/3 if the word is monosyllabic and u_2 is á (exception: gáz) or if u_2 is i (exception: csíz, csiríz)

I/4 if u_2 is a consonant (exception: pénz)

Zs: First Declension:

It is to be mentioned that here we could not find nouns which belonged to I/1. To be examined:

- What is u_2 ?

- is u_3 a palatal or a velar vowel?

I/2 if u_2 is a vowel (exception: rozs)

I/3 if u_2 is a consonant and u_3 is a palatal vowel

I/4 if u_2 is a consonant and u_3 is a velar vowel

C: we found only such nouns with terminal -c as belong to I/3 and I/4. Here several other aspects - unlike the previous ones - had to be taken into account. First of all it is important whether the word is velar or not. If not, it is not enough to start off from its being palatal. The following possible forms can be distinguished:

the vowels occurring in the word are (övä) or
the last vowel of the word is (evévi)

Thus:

I/3 if the last vowel of the word is (evévi)

I/4 if the word is velar (exception: nyolc)
if the vowels of the word are (övä)

Cs: its algorithm is very similar to that of C. Nouns belonging to I/3 and I/4 were found only. If the word is palatal, then we are not given clear-cut information as to the declension-type. Just like with C, (övä) and (evévi) should be distinguished.

Thus:

I/3 if the vowels of the word are (evévi)

I/4 if the word is palatal

if the word is palatal and its last vowels are (övä)
(exception: bölcs)

J: this ending has proved to be the most difficult to classify. The number of new exceptions is considerable as compared to other algorithms. Our "usual" considerations:

- What is u_2 ?
- is the word palatal or velar?
- is the word monosyllabic or polysyllabic?

had to be applied in a combined form in this case. Words with a terminal j belong to I/1, I/2, I/3.

Thus:

I/1 if u_2 is é and the word is palatal
if u_2 is (avávo) and the word is monosyllabic
(exceptions: raj, baj, goj, boy)

I/2 if u_2 is é and the word is velar
if u_2 is (aváve) and the word is polysyllabic

I/3 if u_2 is (ivev consonant)

Ly: Owing to the fact that these letters denote the same sound as j does, in the beginning we intended to draw up a joint algorithm for both of them. Still on the basis of our examples we did not succeed in finding corresponding features in all of the cases.

Words with a terminal ly also belong to the following three types: I/1, I/2, I/3.

To be examined:

- What is u_2 ?
- is the word velar or palatal?

As in the case if j these had to be applied in a combined form.

Thus:

I/1 if u_2 is not (avovö) and the word is palatal

I/2 if u_2 is ö
if u_2 is not (övova) and the word is velar

I/3 if u_2 is (avo); (exceptions: pehely, kehely)

B,P: since these endings exhibit exactly the same peculiarities in their declension, their algorithms can be contracted.

To be examined:

- What is u_2 ?
- is the word palatal or velar?

If the word is palatal, we have the 3d person Singular Possessive Case without a j, but with j if it is velar.

N: This is the first algorithm where it served our aim better to ask for R - the exception chart - at the end. With nouns with a terminal -n we found a lot of exception which were brought about partly by the nature of certain suffixes and partly by other grammatical peculiarities.

Velar nouns with a terminal -n belong - with some exceptions - to the second declension, thus we have a j in 3d person Singular in the Possessive Case. In the determination of declension type here we depend on u_2 .

A noun belongs to

I/1 if u_2 is (evi)

I/2 if u_2 is (övü)

II/2 if u_2 is (avávo), (exceptions: vagyon, rokon, tu-lajdon)

Special mention must be made of the suffix -talan, -telen (-tlan, -tlen). Experience shows that the variant with

j is more frequent in the spoken language (possessive affix in 3d person Singular). In texts taken from literary works, however, the opposite of this prevails. We work with paradigms not containing -j-.

Algorithms of Nouns with Terminal Vowels

We must state that this algorithm was made on the basis of the whole stock of vowels, but we are also aware of the fact that the Telex will not denote vowel every Hungarian. The first step in the algorithm of vowels is to transcribe velars into palatals. It is only i that makes an exception, because it can be both velar and palatal.

The distribution into sub-types takes place according to their endings:

- 1) ő, ü, ű (ó, u, ú)
- 2) i
- 3) e (a)

In order to obtain the desinence of the required vowel harmony with 2 (i.e. nouns with a terminal -i) the vowels of the preceeding syllables have to be subjected to examination. It is only monosyllabic words that bring some difficulties. For simplicity's sake these cases are classed among exceptions.

Type 3) is characterized by a lengthening in the stem before affixes both in palatal and velar variants, denotable by Telex.

It is at the vowels where chart R will be the fullest. Paradigms of words with y in the stem, with -ó, -ő alternating with a/e in the ending, of words vő, fő etc., must be given.

- o -

The algorithms given in the foregoing make it possible to produce any form of a Hungarian noun, but to decide which form is demanded we have to start off on the basis of the information obtained from the Russian analysis.

APPENDIX⁺

Algorithms for determining the type of declension

ε

- 111. u_1 is ε (112)
- 112. Is the word on R (113, 114)
- 113. Take forms from R
- 114. Is u_2 a consonant? (115, 119)
- 115. Is u_2 r? (116, 118)
- 116. Is u_3 (á v ε v i)? (117, 118)
- 117. Turn to I/3
- 118. Turn to I/4
- 119. Is u_2 (a v á v o v u v ö v ü)? (120, 121)
- 120. Turn to I/2.
- 121. Is u_2 (e v é)? (122, 123)
- 122. Turn to I/1
- 123. Is the word of "velar harmony"? (124, 125)
- 124. Turn to I/2
- 125. Turn to I/1

⁺ u_1 = the last letter of the word, u_2 = the second letter of the word from the end, u_3 = the third letter of the word from the end. After each question the first answer is the affirmative one and the second the negative one.

sz

91. u_1 is sz (92)
92. Is u_2 a consonant? (93, 96)
93. Is the word of "velar harmony"? (94, 95)
94. Turn to I/4
95. Turn to I/3
96. Is the word of "velar harmony"? (97, 98)
97. Turn to I/2
98. Turn to I/1

z

61. u_1 is z (62)
62. Is u_2 a consonant? (63, 64)
63. Turn to I/4
64. Is u_2 (e v é)? (65, 66)
65. Turn to I/1
66. Is u_2 á (67, 70)
67. Is the word monosyllabic? (68, 69)
68. Turn to I/3
69. Turn to I/2
70. Is u_2 a ? (71, 72)
71. Turn to I/2
72. Turn to I/3

zs

- 81. u_1 is zs (82)
- 82. Is u_2 a consonant? (83, 86)
- 83. Is u_3 a velar vowel? (84, 85)
- 84. Turn to I/4
- 85. Turn to I/3
- 86. Turn to I/2

c

- 141. u_1 is c (142)
- 142. Is the word on R? (143, 144)
- 143. Take forms from R
- 144. Is the word of "velar harmony"? (145, 146)
- 145. Turn to I/4
- 146. Are the vowels of the word (ö v ü)? (147, 148)
- 147. Turn to I/4
- 148. Turn to I/3

cs

- 161. u_1 is cs (162)
- 162. Is the word in R? (163, 164)
- 163. Take forms from R
- 164. Is the word of "velar harmony"? (165, 166)
- 165. Turn to I/4

166. Is the last vowel of the word (ö v ü)? (167, 168)
167. Turn to I/4
168. Turn to I/3

i

201. u_1 is i (202)
202. Is the word in R? (203, 204)
203. Take forms from R
204. Is u_2 é? (205, 208)
205. Is the word of "velar harmony"? (206, 207)
206. Turn to I/2
207. Turn to I/1
208. Is u_2 (i v e v consonant)? (209, 210)
209. Turn to I/3
210. Is the word monosyllabic? (211, 212)
211. Turn to I/1
212. Turn to I/2

ly

181. u_1 is ly (182)
182. Is the word in R? (183, 184)
183. Take forms from R
184. Is u_2 (a v o)? (185, 186)
185. Turn to I/3
186. Is u_2 ö (187, 188)
187. Turn to I/2

188. Is the word of "velar harmony"? (189, 190)
189. Turn to I/2
190. Turn to I/1

n

41. u_1 is n (42)
42. Is u_2 (e v i)? (43, 44)
43. Turn to I/1
44. Is u_2 (ö v ü)? (45, 46)
45. Turn to I/2
46. Is u_2 (a v á v o)? (47, 48)
47. Turn to II/2
48. Is it in R? (49, 50)
49. Take forms from R
50. Turn to II/2

b v p

21. u_1 is b v p (22)
22. Is u_2 a consonant? (23, 24)
23. Turn to II/4
24. Is the word of "velar harmony"? (25, 26)
25. Turn to II/4
26. Turn to I/3

Vowels

1. u_1 is a vowel (2)
2. Is u_1 i (3, 4)
3. Turn to III/2
4. Transcription: a--e, o--ö, ó--ő, u--ü, ú--ű
5. Is u_1 e? (6, 7)
6. Turn to III/3
7. Is the word in R? (8, 9)
8. Take forms from R
9. Turn to III/1

Algorithm for replacing the palatal endings with the corresponding velar ones after roots of type "velar"

10. Was the word transcribed in 4? (11, 12)
11. The form is right
12. Did it have a final i ? (13, 15)
13. Were there (a v o v u v é) before i ? (14, 15)
14. To be transcribed e--a, ö--o, ő--ó, ü--u, ű--ú
15. The form is right

THE TWO LEVELS OF THE SYNTACTIC ANALYSIS IN THE MECHANICAL ANALYSIS OF RUSSIAN TEXTS³¹

by Gy.Hell

1. Differences between morphology and syntax

In recent years opinion has become unanimous concerning the separation of morphology and syntax: the isolated analysis of the single word belongs to morphology and everything beyond this belongs to syntax [2]. It is true that endings expressing grammatical interconnections already bear a syntactic character, nevertheless it turned out to be suitable to include into the operations of morphological analysis in MT only the operations referring to a string between two gaps.

The objectives of the syntactic analysis are not uniform. The definition that the syntactic analysis has to determine the "boundaries" and the roles of the different elements (groups) of the sentence comprises two different types of problems if we consider the analysis as an analysis from below [3]. The generative analysis subdivides the sentence into two subgroups, the function of which is already determined by the "bisecting". If we consider, however, the analysis as an analysis from below, it is impossible to divide the sentence into two parts with the same lack of sophistication, called "nominal group" (noun phrase) and "verbal group" (verb phrase) resp. This grouping will be reached only at a higher level of the syntactic operations on the basis of a preliminary grouping. If the analysis from below is the opposite of the analysis from above, it seems to be unsophisticated not to begin with a similar arrangement of the word sequences as is usually with the immediate constituent analysis [4].

We have to differentiate, above all, between the two sorts of relations between the words discussed below. The first cor-

respondence consists of linking the words of the sentence with the predicate directly, the other connects the words with a word, called base, which builds up a direct connection with the predicate from the indirect one or which represents directly the indirect connection. The first connection represents the proper syntactical connection, the relations subject-predicate, predicate-object, predicate-adverbial of the traditional grammar. These interconnections depend on the nature of the predicate, they are formed according to the governments of the predicate and their identification comes up against a difficulty at the initial stage of the analysis from below, especially in Russian sentences. This concerns particularly the interconnections between the predicate and the adverb which can only be identified if besides the government of the predicate we take into account also the governments of other words being in the same construction as the given word.

We do not divide the Russian sentence after the morphological analysis is carried out directly into two parts: the nominal part containing the subject and into a verbal part containing the predicate and its objective and adverbial connections. Instead we divide initially the sentence into two parts: one part contains all the nominal structures which can be brought into direct connection with the predicate by means of one or another word but which do not contain in themselves the characteristics of a connection with the predicate. (Moreover, such a structure may be a predicate in itself, too). The other part (the verbal structure) consist only of such elements which are in direct connection with the predicate (also the subject may occur among them).

2. The two levels of the syntactic analysis

The operations of the syntactic analysis on the two levels are based on different aspects. The first level, which we shall call structural level, is based, above all, on the agreement, the second (the proper syntactic level), on the government.

The characteristic features of the analysis on the structural level are as follows:

a) It is built on the rules of agreement of the Russian language, on the use of the comma, on the rules of word order typical of the substantival group.

b) Each word of the sentence is assigned either to a nominal or to a verbal structure.

The analysis on the syntactic level exhibits the following properties:

a) It states connections between the word groups established on the structural level (or between single words of equal value, respectively) on the basis of government and semantic features in a way to obtain structures (or a single structure only) grouped around the predicate of the sentence;

b) it inquires into the quality of the connections between the words of the different groups (with regard to the requirements of the target-language) especially on the basis of the semantic categorization of the words.

Obviously, it is impossible to make any hard and fast line of demarkation between the structural level and the syntactic one. There exist sentences in the analysis of which some questions of the structural level can only be solved by the syntactic analysis. The structural level precedes, as a matter of fact, the syntactic level in MT.

The analysis has always its origin in the given sentence and tries to find out the dependency relations within this sentence and to state the grammatical structure of the sentence. The aim of the grammatical analysis, however, is not to find out whether the general characteristics of Russian sentences may be discovered but to establish a satisfactory sentence structure on the basis of the given data [5].

The data stored in the dictionary give the starting material for the structural level of the mechanical analysis. In the dictionary every word is substituted by the form class category of the stem morpheme and by the grammatical information received by means of the analysis of the ending morphemes. The preposition is not taken for a word of full value which governs the substantive but for a free morpheme which yields with the ending the case category of the substantive [6].

3. Nominal structures with subordinate elements standing before the substantive and being in agreement with it

One of the most important word sequences is grouped in Russian around the substantive. In most cases, the formal properties of interdependence are determined by the agreement.

- a) При таких ядерных реакциях выделяется атомная энергия.
- b) Натрий имеет некоторые типичные металлические свойства,
- c) Солнечная атомная энергия играет исключительно важную роль в нашей жизни.
- d) Многие металлы при очень низких температурах обладают свойством сверхпроводимости.
- e) При слабом протравливании поверхность каждого зерна может быть покрыта сплошь мельчайшими фигурами травления.

The substantive builds a construction based on agreement with the preceding adjective, pronoun, participle or numeral which may be not only a single word but also a word group. These words may be separated from each other by commas, or may be connected with each other by conjunctions.

The adverb is not joined directly with the substantive, only through the adjective. The adverb belongs doubtlessly to the nominal structure, if it is preceded by another adjective belonging to the substantive or by a preposition (b, d). If

the adverb stands at the beginning of the nominal group (c), it does not belong to the substantival group, especially if a verb precedes it. The adverb of the verb may occur behind the verb, too (e). The proper place of the adverbial may only be determined on the syntactic level in such cases.

4. Nominal structures with elements standing after the substantive and being in agreement with the substantive

The words after the substantive (to the right) may also be connected with each other by agreement:

а) ...силя, действующие в направлении возможного скольжения,...

б) ...алгоритм, по возможности равносильный изучаемому процессу управления,...

с) Обычно кибернетикой называют область науки, изучающую процессы управления и контроля в машинах и живых организмах/.

д) Углерод в стали присутствует в виде соединения его с железом, приблизительно отвечающего формуле и называемого цементитом.

е) Эквивалентным называем такие количества энергии разных видов, которые способны дать одно и то же число килограмметров (или эргов) работы.

ф) Пояснения здесь будут даны в той мере, в какой это необходимо.

In our examples a participle (or participles) and an adjective (or adjectives), respectively, are jointed by agreement with the substantive preceding them. The substantive which is the base of the construction may stand either directly before the adjective and participle, respectively, or also at a distance from it.

In these constructions not only the adjective itself or the participle is joined with the substantive but also a more or less expanded word sequence, the beginning of which is given by the comma between the substantive and the word being in agreement with it, and the end of which is given by the comma standing after the word being in agreement with the substantive (of course, not always the first coma) or by the full stop indicating the end of the sentence. These constructions which are called participial or adjectival constructions in traditional grammar, consist sometimes of several parts and may be very lengthy, too (d).

Another type of connection based on agreement is the so-called relative clause which is linked with the substantive similarly to the participial constructions (e.). The constructions of comparison of different value are connected with the substantive in a similar manner (f.).

The boundary of the nominal structure in these word sequences is, on the one hand, the first word which agrees with the substantive and the last word of the participial construction, of the relative clause or of the clause of comparison, respectively, on the other.

5. Words connected with the substantive but which do not agree with it

From the point of view of the mechanical analysis the clarification of the information with which it is possible to formalize the connections not expressed explicitly by the formal means of grammar, plays a primary role. The connection between the adverb and the adjective is - at least in certain cases - evident. In the sentence 4/b the adverbial construction, belongs also to the adjective by means of the word order connections. Since we do not regard the contraction and the government on the level of the structural analysis as formal ru-

les, we may formalize the connection of the mentioned construction and the adjective by means of the following "rule":

a) A new substantive standing after the given substantive which is not in agreement with the latter does not terminate the nominal structure if an element follows it, separated by a comma, which agrees with the first substantive.

b) Such a word sequence is closed, between its boundaries the elements belong together and the first substantive is the real (or fictive) base of the sequence [7].

On the basis of this rule (see sentences 4/c and 4/d) it follows, that the sequences *область науки* and *соединения его с железом* also form a formal connection with each other. We find an example for a similar connection in sentence 4/e, too: *такие количества энергии разных видов*.

According to the above we may make the following statements:

1) We must take for granted the possibility of a direct connection between two substantives following each other but which are not separated by a comma from each other, if the second substantive is in genitive or in a subordinate case or preceded by a preposition (we have considered the preposition as having the same value as suffixes).

2) If such a connection occurs, we must take for granted the formal designation of it.

The connection between the substantives may be extended, because of the closed character of the nominal structures, to other cases, too:

a) Действие других материальных тел на данную материальную точку, результатом которого является ускорение в движении точки, называется силой.

b) Так как освещение поверхности микрошлифа, при рассмотрении его под микроскопом, никогда не бывает строго вертикальными,...

с) ...при этом необходимо учитывать распределение и объем памяти для численного материала и программы, невозможность проведения операции над несколькими числами без фиксации промежуточных результатов в некоторых ячейках памяти, запрет записи нового результата на место нужной в дальнейшем величины и т.д.

д) Управление социалистическим производством и его планирование обеспечивают...

According to the examples we must take the connection of two substantives which are in different cases or in genitive which follow each other and are not separated from each other by means of a comma, for formally designated because of the continuity of the nominal structure, if they are in an intermediary position or form a separated construction.

A substantive is in an intermediary position (or another word of the same value), if it stands between the qualified word and the following participial, adjectival construction, or construction of comparison or enumeration (5/d, see also 10.), i.e. between the substantive and its subordinate clause.

The separated construction is a nominal structure flanked by two commas (3/b) at the very beginning of which stands a rather confinable prepositional group. In such constructions participles or verbal elements cannot occur.

The rule concerning the connections between the substantives in an intermediary or separated position does not tell anything about the way the connection comes into being, by no means in a chain-like connection (5/a, see also 10.) under any circumstances.

6. Particular rules concerning agreement

Some constructions based on agreement are formed in a slightly different manner from the general rules:

а) Как к управляющему, так и к управляемому устройствам может поступать...

б) Простые вещества образуют две больших группы тел, цещащие...

с) Рассмотрим, каковы средства и методы кибернетики, т.е. каковы тот математический аппарат и та конкретная проблематика, с помощью которых...

7. Infinitive in the nominal structure

The infinitive does not show any agreement with the word it is connected with. Its belonging to the substantive may only be shown on the level of the structural analysis by means of the closed character of the nominal construction.

а) Второе соображение, заставляющее начинать курс механики со статики, ...

б) При использовании возможности определить величину методом инверсии, нам дается...

In the sequence а) the infinitive belongs only indirectly to the substantival base. We often come across such positions of the infinitive in technical texts. The example б) shows the case when the infinitive is in a closed position and may only belong to one of the preceding substantives (with greatest probability to the immediately foregoing one).

The connection of the infinitive with the base may be determined comparatively rarely after the first step of the syntactic analysis.

8. The structural formula of the nominal construction

The possible types of substantival structures (NOM) may be summarized in the following formula [8]:

$$NOM = N (+N') \left\{ \begin{array}{l} \left[+ K \left\langle (+ D +) \left\{ \begin{array}{l} PA \\ AA \end{array} \right\} + \left\{ \begin{array}{l} Pn \\ N (+N') \\ INF \\ D \end{array} \right\} \right\rangle + K \right] \\ \left[+ K \left\langle (+ N +) C + VER (+ NOM) \right\rangle + K \right] \end{array} \right\}$$

It seems to be expedient to write the separated structures ('NOM') in a special formula:

$$'NOM' = \left\langle K + N (+N') (+INF) + K \right\rangle$$

We have here:

INF = Vi + NOM

N = (AD +) Su

Su = (Pe +) S

AD = $\left\langle (Pn+) (Nu+) (D+) (A +) (Pa +) (K +) \right\rangle$

S = substantive

Vi = infinitive

Pe = preposition

Pn = pronoun

A = AA = adjective

Nu = numeral

Pa = PA = participle

D = adverb

K = comma, copulative or comparative conjunction (perhaps: full stop)

C = relative pronoun or a corresponding conjunction

(); $\langle \rangle$; [] = possibility

{ } = alternative possibility

We distinguish in the formula between two kinds of participles and adjectives:

PA and AA stand on the right of the substantive and are separated from it by means of a comma and are in agreement with it (PA = participle, AA = adjective), continuing the substantival structure at the same time.

With suitable indexes the different elements of the construction may be placed in the right places. The N without any index denotes the base of the structure. (Further, only the elements S of the construction N will get indexes).

N' denotes the group which is connected with another preceding construction N without any agreement. (The conditions for such a connection are given in the formula). Several groups N' following each other do not show any differences in indexes, so we denote, for instance, in the sentence 4/f both энергия and видов with S'.

The element of the group in the brackets after PA, AA and C, respectively (PN, N, INF and D) must have the symbols PA, AA or C, respectively. So, for instance, "в направлении" the sentence 4/a has the symbols N-PA and Pe + S-PA, respectively; "возможного скольжения" has only the symbol N', because we must only indicate, that the adjective + substantive belongs to the preceding N.

9. Irregular constructions in nominal groups

The arrangement of the nominal structure according to the above formula may be called "regular" construction. A great part of nominal structures fits into this scheme, so the deviations can be systematized much easier:

a) Все эти виды делим на два класса: энергия кинетическая и энергия потенциальная.

б) ... в основе ее лежит представление ...

с) ... треугольные, квадратные или другой, менее правильной формы углубления...

д) ... постепенно увеличивающиеся при растяжении напряжения..

е) ... результаты, для наглядной иллюстрации на графике представленные, ...

In the first group some elements of the group AD may stand also after the substantive.

$N (+AD')$

To this group AD' adjectives and pronouns may belong (e.g. a) and b)).

Among the substantives standing on the right of the given substantive and being connected with it by a genitive we distinguish a subgroup (N'') which may also precede the substantive which represents the base of the construction (c.). Such substantives have always an element AD [9].

$(N'') = (AD+S)_G$

The construction 9/a occurs only rarely, the greatest part of the anomalous constructions is contained by 9/d. This is, as a matter of fact, a construction PA (or AA), which precedes the given substantive. We denote them in the formula with PA'' and AA'' , respectively.

If we denote the construction N complemented in this way with NO, we get

$$NO = \left\{ \begin{array}{l} PA'' + N + \\ AA'' + N + \\ N'' + \end{array} \right\} N (+ AD'')$$

According to the new denotation we have to transform the formula of the substantival construction already known. We substitute NO for N:

$$\text{NOM} = \text{NO} (+\text{NO}') \left\{ \begin{array}{l} \left[+K \langle (+D+) \left\{ \begin{array}{l} \text{PA} \\ \text{AA} \end{array} \right\} \left\{ \begin{array}{l} \text{Pn} \\ \text{NO} (+\text{NO}') \\ \text{INF} \\ \text{D} \end{array} \right\} \rangle + K \right] \\ \left[+K \langle (+N+) \text{C} + \text{VER} (+\text{NOM}) \rangle + K \right] \end{array} \right\}$$

$$\text{"NOM"} = \langle K + \text{NO} (+\text{NO}') (+\text{INF}) + K \rangle$$

The construction PA" and AA" are essentially identical with the constructions PA and AA, respectively. For the sake of distinction we may call the latter open constructions and the former closed ones. We do not consider these constructions as a subgroup of the so-called prepositional constructions in which the elements are connected by means of strong and weak government. The formal character of the connection means that the participle or the adjective attaches to the substantive some other additional words by agreement, which are terminated in the case of the construction PA" (or AA") by the substantive designated by the agreement and in the case of the construction PA (or AA) by the comma.

The degree of the formal definability enables us to distinguish the constructions PA" and AA" from the constructions (+NO'). If we consider also the government as a fundamental operation of the level of the structural analysis, the two constructions will not differ from each other essentially, which however leads, to difficulties in making up the algorithms.

10. Nominal structures with regular and inhomogenous word order

A very frequent case of the formal connection based on agreement is the enumeration of nominal structures of identical case.

a) ... как в жидких, так и в твердых растворах, и, в частности, в металлических сплавах ...

b) ... сочетать быстроту синтеза, универсальность метода и его экономичность ...

c) ... исходная управляющая система, обменивающаяся информацией с действующим устройством и работающая согласно некоторому алгоритму, который учитывает ...

d) Большое значение для техники имеет применение алюминия, металла, наиболее распространенного в земной коре.

e) Планирование включает установление структуры, темпов и пропорций развития социалистического производства, определение заданий по качеству его работы, а также путем выполнения плановых заданий.

f) Отпуск заключается в нагреве закаленной стали до невысокой температуры, определенной выдержке и в охлаждении.

g) Разница в массах изотопов объясняется различным содержанием в ядре нейтронов - нейтральных частиц, с массой, равной приблизительно массе протона.

Different nominal structures - complete and incomplete ones - may form a construction: 1) the adjectives standing before the substantive, the pronouns standings before the substantive, etc.; 2) the substantives representing the base of the nominal group (10/b, 10/f,); 3) participial constructions (joined with the participle PA), (10/d); 4) substantival groups (NO') joined not by agreement with the substantive e.g. d, e, g; 5) sentence-like groups joined by a relative pronoun with the substantive, structures of comparison.

The groups are joined with each other by comas and by a particular group of conjunctions. The commas and conjunctions are supplied with different place values according to the quality of the joined elements (NO, NO', N-PA, etc.).

In the case of constructions of enumeration the first member of the enumeration is the fictive base of the others.

We distinguish in the connection of the two substantival groups denoted by NO and NO', respectively, two subgroups, which may be separated easily with the help of the rules of punctuation in Russian. The parting of нейтронов - нейтральных частиц in the sentence 10/g by a blank indicates a particular connection between the members of the enumeration. It is possible to connect the words структуры, темпов of the sentence 10/d in the same manner, but we cannot put a blank for instance between алюминия - металла of the sentence 10/e.

The way of the connection of the constructions of enumeration, the applied conjunctions and the punctuation marks play an important role in the analysis of the sentence.

If we wish to express in the already used formula the abovementioned constructions of enumeration, we need not make substantial alterations. We must only allow the possibility of the iteration of all of the connected parts or subparts of the group. We denote the possibility of the iteration by an asterisk but for the sake of simplicity only with the symbols S, K, C, PA, AA, PA" and VER.

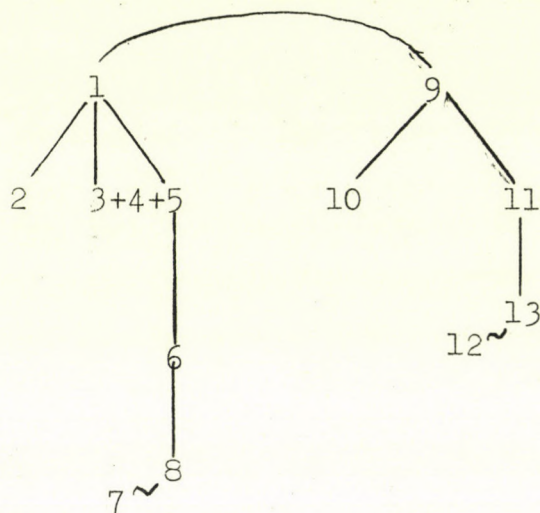
According to the above we may assign the following sequence of indexes to the sentence 10/b:

$$\dots V + S + S\text{-NO}' + K + S^x + S\text{-NO}' + C + Pn + S^h \dots$$

It is even possible to express the more complicated relations of the part (установление ... работы) in the sentence 10/e with the help of this notation:

$$\begin{array}{cccccccc}
 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
 \dots & S + S-NO' + K + S^x-NO' + C + S^x-NO' + S-NO' + A + S-NO' + \\
 & 9 & 10 & 11 & 12 & 13 \\
 + K + S^x + S-NO' + Pe + S^x-NO' + Pn + S-NO' + K. + \dots
 \end{array}$$

If we represent the dependence of the members of the nominal construction of each other, we have the following diagram:



Before the more complete description of the word sequence 10/e we have to mention another type of the constructions of enumeration, namely when the enumerated element is not of basic type. Two such NO'-groups, although belonging to the same base NO, are not separated from each other by a comma or a conjunction. E.g.

h) ... применение при его выполнении общепринятых методов ..

In the above word sequence two NO'-elements are connected with the substantive применение. (We leave now aside the different nature of the joining, on the one hand, because we are only interested at present in the fact of the connection, and S contains also the case of the word, so also something of the nature of the connection, on the other.) According to the above the following formula expresses the connection satisfactorily:

$$S + Pe + Pn + S-NO' + A + S^x-NO'$$

We must also consider that such a connection is not only possible between substantives following each other but also between substantives standing quite at a distance from each other. Thus, we apply a particular index both on the base and the dependent member of the connection:

$$S_{\alpha} + Pe + Pn + S-NO' A + S_{\alpha}^x-NO'$$

The mentioned example poses also the question in which way several NO'-elements may be attached without agreement to one NO-member. With respect to Russian it is quite impossible to state as a general rule, that the NO'-element in genitive stands always after its base [11]. Also in the example 10/h the order is reversed:

$$NO + S_{(Pr)}-NO' + S_{(B)}^x-NO'.$$

The rule of order with respect to the nominal structures in Russian may be formulated in the following manner:

1) Substantives in genitive without any preposition and following each other are joined chain-like to a sequence of possessives;

2) if there is a substantive not in genitive in the sequence of the NO'-elements, so

a) the member which is not in the genitive is by no means the NO'-member joined with the preceding element,

b) the possessive substantive following it is not always the NO'-group of the preceding member.

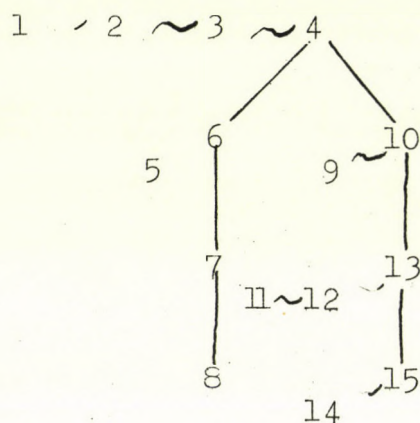
E.g. i) Эта формула представляет собой наиболее общее математическое выражение физического закона сохранения массы для рассматриваемого случая плоского одномерного движения сплошной среды.

According to our notation:

1 2 3 4 5 6 7 8 9
 D + A + A + S + A + S-NO' + S-NO' + S-NO' + Pe + Pa +

10 11 12 13 14 15
 + S^x-NO' + A + A + S-NO' + A + S-NO'

The diagram of the sequence:



It seems to be worth while to return to the sentence 10/e once again, because it is extraordinarily interesting for some reasons. The путем выполнения плановых заданий is joined with the preceding member of the construction of enumeration by the conjunction а также. In the foregoing examples the connection consisted of joining groups of the same case which we considered as units for the same reason. The analysis of sentence 10/3 shows that different prepositional substantives may also occur as members of equal value within a sequence, although their identity is not a morphological one. In this way both по качеству and путем выполнения are identical.

In example 10/c two participial construction are attached to the base система; still a constructions K + C + VER (+NOM) is joined with the second participial construction. The existence of any connection can be proved also by formal means, from which it follows that the different dependent constructions in the

nominal structure may not only be iterated in the nature of enumeration but also may join with each other by alternation.

The situation is different if the conjunction C' cannot exhibit a formal agreement (i.e. it is not joined with one substantive of the nominal structure but with a verbal member of dependent position).

j) Символом $p(x \sim y)$ обозначают логическое условие, выполненное в том случае, когда неравенство, стоящее в скобках, истинно.

Such constructions are to be taken - similarly to the infinitive constructions - for ambiguous.

11. Elliptic constructions, nominalization

a) При этом необходимо учитывать...

b) Мы должны принять как доказанное, что ...

c) Последний из работающих операторов содержит ...

d) Из-за этого стола нет в комнате.

e) ... назовем равнодействующей данных сил.

f) На производства объяснения и указания рабочему, выполняющему работу, почти всегда даются с помощью чертежа.

g) ... под действие одной силы, равной сумме данных сил.

h) Такое решение оказывается возможным.

With the help of the regularities being manifested in the construction of the nominal structure it is quite possible to clear up - without any references to the vocabulary - the substantival role of the members P_n , A and P_a which play this

role instead of the failing member S of the sequence NO. In the case of a), b) and c) 3 the nominalization results from the fact that the group NO cannot occur without S.

In one group of the elliptic constructions (in the constructions which are similar to c)) - by virtue of a particular type of agreement - it may be set down as a fact that the substantivized Pn, A and Nu form one unit with the directly following NO'-group. For the analysis one has to start from the connection of the member which are denoted with special indexes in the vocabulary:

$$\langle (Pn'') (A'') (Nu'') + из + S_{plG} \rangle$$

In the case of the sentence e) everything is present which is needed for a NO-group. The requirements of agreement allow, however, only the following construction:

$$V + S + Pa + S-NO'$$

where: S = Pa.

For the purposes of a complete analysis of the sequences d), f) and g) we have to make use also of the methods applied on the syntactic level. In f) ... рабочему, выполняющему работу, ...

enables neither the sequence A + K + Pa + S + K

nor the sequence A + K + PA'' + S + K

(in the first case it is the agreement, in the second the construction PA'' which is decisive). The requirements of the nominal structure are only fulfilled in a satisfactory way by the sequence

$$S + K + PA + S + K$$

About the syntagma равной суммы of the example g) we may only affirm that it is of value AA + S - AA (and not

of A + S), if in the syntactic analysis it has already turned out that this part is a direct continuation of the preceding one.

The adjective *возможный* in the sequence h) is similarly of uncertain position.

12. The construction of verbal structures

For the determination of the verbal structures it is already far more difficult to find out a connection of general validity which may also be described with the help of formal categories than in the case of nominal groups. Even the name "structure" is not correct to a certain extent: according to this name none of the verbal groups could do without a verb (in infinite or finite form) but a great deal of Russian sentences fail to have a verb at all.

a) Мы хотим данный процесс управления, т.е. данный процесс переработки информации, вложить...

b) Твердый цементит можно легко отличить под микроскопом, сделав предварительно отпечатки алмазной пирамидкой, вдавливающейся в образец под определенной нагрузкой.

c) Поэтому естественно после принципов механики начать обзор содержания ...

d) Необходимо уметь измерить.

Between the members of the verbal group there does not exist that feature of agreement which played such an important role in clearing up the nominal structures. The verbal groups fail to have also the continuity and the members of the group may stand rather far away from each other (12/a).

The verbal structure is not completely separable from the nominal structure (the connection was also to be seen in the

formula of the nominal structure) and the nominal structures may occur as groups depending on the independent verbal parts in the verbal structures (12/b).

The connection between the two structures comes to light both in the formula of the nominal and in that of the verbal group: on the side of the determination also the other characteristic group may occur but the member to be determined itself, too. In this way we have theoretically the possibility of infinite connections which of course does not occur in reality. (We must not overlook the fact either that the identical member on both sides of the formula are only seemingly identical: with the independent or basic member some members of the same type are joined.)

The central part of the verbal group is represented by the independent finite and infinite verb forms, the short adjectives, the participles and infinitives and the different adverbs; they represent, therefore, the base of the verbal structure 12. The present participles (participial groups) some particular groups beginning with certain conjunctions (here we must take into consideration also the type of the sentence, i.e. whether it is interrogative, affirmative or imperative) and perhaps a group introduced by an infinite verb form are connected with the central part. In the verbal structure the different members and subgroups may repeat themselves in the same manner and be connected with each other as in a nominal structure.

The formula of the verbal structure:

$$VER = \left\{ \begin{array}{l} V \\ Vau + Inf \\ P\bar{a} \\ A^- \\ (D) \end{array} \right\} \cdot (+D) + \left\{ \begin{array}{l} G < (+D) (+Inf) (+NOM) > \\ K + C'' + < (+VER) (+NOM) > \\ Inf (+NOM) \end{array} \right\}$$

independent members

Here too we introduce indexes to distinguish the members of the structure: the members of the first part are not given any index, every member of the second part gets the indexes G, C" and I, respectively.

13. The syntactic structures of the sentence

Each sentence consists of two large structures:

$$S = (\text{MOM} +) (+ \text{VER})$$

These two groups are not the subject group and the predicate group of the sentence. Even if we take no notice of the fact that there are a great many sentences in Russian without verbal structures, the obtained subparts have nothing to do with the categories called the main parts of the sentence. The sentence is divided into two parts in order to get a base for connecting the words (syntagmas) of the sentence in the further analysis, the result of which will be really the determination of the subject group and the predicate group, respectively. The new division happens on the second level of the syntactic analysis, in the proper syntactic analysis.

We shall illustrate the structural analysis on the following example:

Многочисленные опыты по исследованию атомного ядра выдвигали вперед важную проблему ядерных сил, необычные свойства которых представляли неразрешимую задачу.

Nominal structures:

- 1) многочисленные опыты
- 2) по исследованию атомного ядра
- 3) важные проблемы ядерных сил, необычные свойства которых представляли неразрешимую задачу

Verbal structures:

1) выдвигали вперед

The proper syntactic analysis concerns only the bases and as a result of this we obtain the division into VP which reflects the relations of the sentence-elements.

14. The determination of the predicate in the syntactic analysis

Further, we shall only examine the most important features of the syntactic analysis.

No doubt, the central question of the syntactic analysis is the determination of the predicate. After the analysis on the structural level we receive the predicate from the verbal group, if any.

The independent members of the verbal structure:

Inf	- infinitive (main verb)
Va,b	- continuous verb: present (a) - past (b) perfective verb: future (a) - past (b)
Vau-1i	- быть, бывать
Vau-1a	- есть, бывает, бывают
Vau-1c	- будет, будут
Vau-1b	- был, бывал
Vau-2i	- стать, служить, оказывать(ся), казаться ...
Vau-2a	- станет, станут, служит, служат, оказывает(ся) ...
Vau-2b	- стал, служил, оказывал(ся), ...
Vau-3a	- могу, можешь ..., хочу, хочешь ...
Vau-3b	- мог, могла ..., хотел, -ла ...
Vau-4a	- приходится, следует
Vau-4b	- приходился, следовал ...
VD-1	- можно, надо, нужно, необходимо
VA-1	- должен, а-на..., способен, принужден ...
Vau-5i	- представлять(ся собой), являться
Vau-5a	- представляет(ся собой) ..., является ...
Vau-5b	- представлял(ся собой) ..., являлся ...

- $\bar{P}a-2$ - short form of the past passive participle
 $\bar{P}a-1$ - short form of the present passive participle
 $P-1$ - бы
 $P-C$ - чтобы
 $P-2$ - нет
 $A\bar{a}$ - short adjective, predicative interrogative words, adverbs

The following combination may give the predicate of the sentence on the basis of the verbal structure:

A, combinations with one single member

- | | |
|-----------------|--------------------|
| 1: Va | 6: $A\bar{a}$ |
| 2: Vb | 7: $P-2$ |
| 3: $VD-1$ | 8: $Vau-1a (+NO)$ |
| 4: $\bar{P}a-2$ | 9: $Vau-1b (+NO)$ |
| 5: $\bar{P}a-1$ | 10: $Vau-1c (+NO)$ |

B, combinations with two members:

- | | |
|--|---|
| 1: $Vau-1c + Inf.$ | 3: $Vau-1b + \bar{P}a-2$ |
| 2: $Vau-1c + \bar{P}a-2$ | |
| 4: $\left\{ \begin{array}{l} Vau-3a \\ Vau-3b \\ Vau-4a \\ Vau-4b \\ VD-1 \\ VA-1 \end{array} \right\} + Inf$ | 6: $\left\{ \begin{array}{l} Vau-2a \\ Vau-2b \\ Vau-5a \\ Vau-5b \end{array} \right\} + NOM$ |
| | 7: $\left\{ \begin{array}{l} Vb \\ Vau-1b \end{array} \right\} + P-1$ |
| 5: $\left\{ \begin{array}{l} Vau-3 \\ Vau-3b \\ Vau-4a \\ Vau-4b \\ VD-1 \\ VA-1 \end{array} \right\} + Vau-1i (+MOM)$ | 8: $P-C + \left\{ \begin{array}{l} Inf \\ Vau-1b \end{array} \right\}$ |

C, combinations with three members:

$$1: \left\{ \begin{array}{l} \text{Vau-3a} \\ \text{Vau-3b} \\ \text{Vau-4a} \\ \text{Vau-4b} \\ \text{VD-1} \\ \text{VA-1} \end{array} \right\} + \text{Vau-2i} + \text{NOM}$$

$$2: \text{P-C} + \left\{ \begin{array}{l} \text{Vau-1b} \\ \text{Vau-2i} \\ \text{Vau-5i} \\ \text{Vau-5b} \end{array} \right\} + \text{NOM}$$

$$3: \text{VD-1} + \text{Vau-1c} + \text{Inf}$$

$$4: \text{VD-1} + \text{Vau-1b} + \text{Inf}$$

$$5: \left\{ \begin{array}{l} \text{Vau-3b} \\ \text{Vau-4b} \end{array} \right\} + \text{P-1} + \left\{ \begin{array}{l} \text{Inf} \\ \text{Vau-1i} \end{array} \right\}$$

D, combinations with four members:

$$1: \left\{ \begin{array}{l} \text{VD-1} \\ \text{VA-1} \end{array} \right\} + \text{Vau-1b} + \text{P-1} + \text{Inf}$$

$$2: \text{P-C} + \left\{ \begin{array}{l} \text{VD-1} \\ \text{VA-1} \end{array} \right\} + \text{Vau-1b} + \text{Inf}$$

$$3: \text{P-C} + \text{Vau-3b} + \left\{ \begin{array}{l} \text{Vau-1i} \\ \text{Vau-2i} \end{array} \right\} + \text{NOM}$$

During the analysis of the Russian sentence it occurs frequently that we do not find any verbal structure. For MT the boundaries of the sentence are given, within which also a predicate must occur (the so-called defective sentences occupy a special place). In ten sentences consisting

only of nominal structures the predicate has to be chosen from these structures on the strength of the word order rules.

15. The objectives of the syntactic analysis in the determination of the sentence-elements connected with the predicate

In the case of a predicate received from a verbal structure the subject of the sentence may be determined on the basis of the agreement rules of subject-predicate. If there does not exist in the sentence such a nominal structure which can act as a subject of the sentence, so the subject is to be found in the verbal structure. The latter may be a single infinitive or a dependent group (denoted by C") of the verbal structure, e.g. При определении размеров получилось, что величина угла ...

Up to the present we could not find a satisfactory solution with the help of the methods of the formal analysis for the purposes of MT for the case when such a subject must be distinguished from the object having the same form. Not even the interpretation of the words whether they can act as a subject or not turned out to be enough.

The connection of the object and the adverb (or adverbs) of the sentence may be carried out on the basis of vocabulary information giving the government of the verbal part.

In the syntactic analysis especially the determination of the adverbial complements raises some difficulties namely the adverbs expressed by prepositional substantives do not differ, from the adjective expressed by an prepositional substantive which is not in agreement with the substantive [13]. A lot of people have dealt with the analysis of these grammatically homonymic constructions but without any satisfactory solution [14]. From the point of view of the translation into Hungarian this question is therefore of primary importance because in Hungarian these adverbial constructions and attributes of adverbial form

show only rarely a total formal agreement, in most cases they require a particular translation. We have to find a way of analysis which diminishes already at a preliminary stage of the syntactic analysis the number of such ambiguous words and syntagmas. The elimination of ambiguities on the basis of the substantival governments does not seem to be a satisfactory solution [15].

16. The analysis of subordinate clauses

On the level of the structural analysis (if there are no uncertain members) the subordinate clauses do not act as a separate group, only as dependent members of a substantival or a verbal base. The structural analysis reduces the compound clauses to a single principal clause and carries out the syntactic analysis only within this sentence, i.e. between the bases of the nominal structure and the predicate (perhaps including the dependent groups of the verbal structure). The analysis of the subordinate clauses will begin only after the analysis of the principal clause had been carried out in a satisfactory manner.

The subordinate clause appears after the structural analysis as a nominal structure connected with the base which has, however, on his part also a nominal and a verbal part.

The analysis within the subordinate clause is carried out in a similar manner as in the principal clause: it means the clearing up of the relations predicate-subject, predicate-object and predicate-adverb.

N O T E S

- 1) As a result of the mechanical analysis we wish to obtain the grammatical interconnections of the words of a given sentence being correct also from the point of view of the "content" of the sentence.

This requirement is an ideal one and the algorithms may only aim at an approximative elimination of grammatical ambiguities.

The interpretation of the semantic unit see in *Studia Grammatica* I, 5.3. p.25.

- 2) The MT-conference in Denver (1962) reflects the same opinion (Cf. Summary, pp.41-60). Mel'chuk sets apart morphology from syntax in the same manner. (МП и прикл.лингв. p.51.)
- 3) This definition is from Garvin (Syntactic Retrieval 287.).
- 4) See on the relation of the two models: Ревзин: Модели языка. p.12.
- 5) See on the impossibility of sentence structures of general validity: Berrár, J., The formal properties of the sentence, in Hungarian, p.73.
- 6) The same interpretation of the case categories may be found in Hjelmslev (Le catégorie de cas, p.105).
- 7) This statement is only correct with some restrictions for other Indoeuropaen languages for e.g. the attributive clause does not always join continuously, (Harris, String Analysis, p.14 and Bierwisch, Syntaktische Struktur, p.259.).
- 8) This structural formula of the nominal structure (even it in not so contracted a form) is to be found also in the transformational grammars. According to the opinion of the representatives of transformational grammar the identification of them is possible, however, only with the help of transformations, see Bierwisch, Syntaktische Struktur, p.286.

- 9) Their analysis merits particular attention from the point of view of the translation from Russian into Hungarian because they cannot be translated by a possessive relation.
- 10) See Summary, p.38.
- 11) With respect to the nominal structure of German see Bierwisch: Syntaktische Struktur, p.259.
- 12) With respect to English the words directly connected with the verb (within the principal clause) are represented in a similar manner by A.K.Joshi (Computation, p.835) and W.Plath (Automatic, p.180)
- 13) With Harris one group of the right adjuncts of the substantive (r_n 1-9) agrees with one group of the objective adjuncts of the verb (α 33.1-9), (String Analysis, p.36)
- 14) Still in the early years of MT L.Brandwood dealt with similar constructions in German (Mechanical Resolution, p.140). With respect to Russian Selimova gives a detailed analysis.
- 15) Sipőczy dealt with the relations of the substantival and Verbal governments in Вопросы программирования русских предложных оборотов. Slavica, 1963, pp.59-72.

BIBLIOGRAPHY

- [1] BÉRRÁR, J., The Formal Properties of the Sentence, (in Hungarian), in Általános Nyelvészeti Tanulmányok I., pp.53-76, Publishing House of the Hungarian Academy of Sciences, Budapest 1963.
- [2] BIERWISCH, M., Syntaktische Struktur deutscher Nominalgruppen, in Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 14, pp.244-278, 1961.
- [3] BOTH, A.D. - BRANDWOOD, L. - CLEAVE, J.P., Mechanical Resolution of Linguistic Problems, London 1958.
- [4] GARWIN, P.L., Syntactic Retrieval, in Proceedings of the National Symposium on Machine Translation, pp.286-292. Prentice-Hall, 1961.
- [5] HARRIS, Z.S., String Analysis of Sentence Structure, Mouton and Co., 1962.
- [6] HJELMSLEV, L., La catégorie des cas. Étude de grammaire générale, Acta Jutlandica, 1935
- [7] JOSHI, A.K., Computation of Syntactic Structure, in Information Retrieval and Machine Translation, pp. 831-840, New York, 1962
- [8] МЕЛЬЧУК, И.А., Об алгоритме синтаксического анализа языковых текстов (общие принципы и некоторые итоги), in Машинный перевод и прикладная лингвистика 7, pp.45-87, Москва 1962
- [9] PLATH, W., Automatic Sentence Diagramming, in Proceedings of the 1961 International Conference on MT, vol.I., pp.175-194, 1962 (Her Majesty's Stationery Office)

- [10] РЕВЗИН, И.И., Модели языка, Москва, 1962
- [11] ШЕЛИМОВА, И.Н., Установление синтаксических связей предло-
жных групп в русском языке, in Лингвистические
исследования по МП. 2, pp.117-141, ВИНТИ,
Москва, 1961
- [12] ШИПЕЦИ, Д., Вопросы программирования русских предложных
оборотов, in Slavica III., pp.59-72, 1963
- [13] Studia Grammatica I-II, Berlin 1962-63
- [14] Summary of the Proceedings of the Conference of Federally
Sponsored Machine Translation Groups on MT-
oriented Syntactic Analysis, Wayne State
University, 1962

SOME SEMANTIC, ASPECTS OF THE MACHINE TRANSLATION FROM RUSSIAN INTO HUNGARIAN

by

Gy. Sipőczy

1. The morphological analysis has to determine the grammatical role of certain morphemes (concerning tense, case, etc.) in a given sentence. In some cases, however, the grammatical role of the morpheme cannot be found out only with the help of analysis of the word form under consideration. In such cases the analysis must also bear on the context. We may find, for instance, a lot of homographs among the substantival morphemes. The question which of the possible cases the given morpheme really represents in the analyzed sentence may usually be answered by the aid of the analysis of the foregoing adjectives and prepositions which are in agreement with the given word. E.g. точки = singular genitive, plural nominative and plural accusative; but на эти точки = plural accusative. Seeing that in such cases the analysis of the prepositions and the substantive was necessitated by the uncertainty of the given substantival morpheme, we may - with good reason - consider the analysis of the whole syntagma (nominal group) as a part of the morphological analysis.

If the role of a homographic morpheme cannot be cleared up by the aid of the analysis of other morphemes belonging to the same syntagma (either because of the presence of another homograph word form or because of the lack of further morphemes in the syntagma at all), an additional - syntactic or semantic - analysis is needed [5, 7, 8].

The analysis of verbal forms consisting of several morphemes also belongs - in the case of compound tenses - to the sphere of morphological analysis. The same holds true if we are to analyze nominal predicates, connected with verbs as быть,

являться, представлять собой. E.g. будет выражать; был создан; был предметом; является предметом; представляет собой машину.

In the case of syntagmas of the type "нужно производить, может быть выражен" it may be questionable - because of the modifying role of "нужно, можно" - whether their analysis is to be considered as a morphological or as a syntactical one. This question is, however, a purely theoretical one and has no great importance from the point of view of the practical work in MT. But it seems more suitable - because of the delimitation of the morphological analysis from the syntactical one - to consider the analysis of syntagmas of the abovementioned type as a part of the morphological analysis, too.

In this way, morphological analysis includes the analysis of separated morphemes, of so-called attributive constructions in which attributes are in agreement with substantives and the analysis of predicative constructions.*

2. The syntactic analysis has to clear up the role of the morphemes (sequences of morphemes) in the sentence, i.e. the interconnections among the different words (word-groups) of the sentence under consideration. As the results of that analysis we must obtain the structural formula, on the basis of which the corresponding sentence in the target-language may be built up. The syntactical analysis necessitates such a code-system which gives suitable information with regard to the matter, with what kind of elements each element of the sentence under consideration (the morphemes or morpheme-groups of the sentence) may form a syntagma, i.e. what kind of "government" they have. (For the present purposes we shall understand

*/ Work on MT in Hungary has so far included only morphological analysis of Russian. (Cf. Computational Linguistics 1, Edited by the Computing Centre of the Hungarian Academy of Sciences, Budapest, 1963.)

here by "government" all the rules concerning the interconnections between morphemes.)

We shall take the verb as the starting point of the syntactic analysis. The choice of the verb is motivated by the fact that, on the one hand, the verb represents one of the morphologically most identifiable elements of the sentence and that most of the substantival morphemes (nominal groups) are linked with the verb, on the other.

In Russian the morphemes linked with the verb may be divided into two groups:

- a) Nominal complements expressed by cases only (according to the Russian grammatical terminology: подлежащее; прямое и косвенные дополнения,
- b) Prepositional constructions or adverbial complements.

This subdivision is necessitated by practical purposes. The complements expressed by mere cases may be linked with particular verbs, i.e. in the cases of knowing the verb it may be stated which complement may be expected to occur in the given sentence and which cannot occur at all. At the same time the complements of the type b) may occur more or less in connection with each verb (regardless of the meaning of the verb).

A similar subdivision may be found in Tesnière's theory of sentencehood. He calls the complements a) "actant" and the complements b) "circonstant".*/

*/ No doubt, some of Tesnière's ideas are theoretically very disputable (e.g., that the subject is always subordinated to the predicate), but his theory - as S.Károlyi has recently pointed out - contains from a practical point of view many useful points. The subdivision of the complements of the verbs in "actant" and "circonstant" may be a useful basis for the compiling of a suitable code-system for the syntactic analysis in MT. Otherwise, to make a difference between the complements of the verb expressed by mere cases and those expressed by prepositional constructions cannot be considered as an arbitrary procedure only: it is the registration of the regularities concerning the case system of the language.

2.1 With regard to the complements expressed by mere cases we may subdivide the verbs in the following manner:



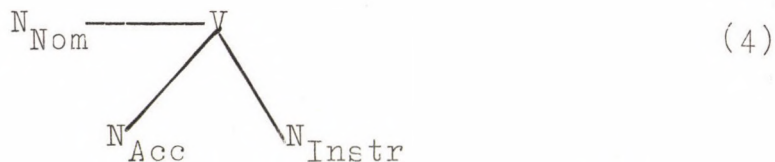
this group contains verbs having only a substantive in nominative (or a nominal group containing a substantive in nominative). This represents for the most part, the group of the intransitive verbs.



i.e. besides the subject the verb governs an objective complement (besides the subject a substantive in accusative or a nominal group containing a substantive in accusative is to be expected).^{*}/



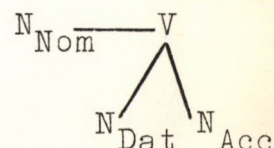
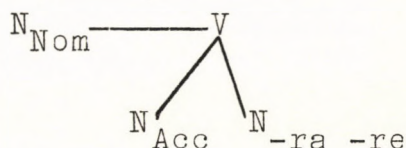
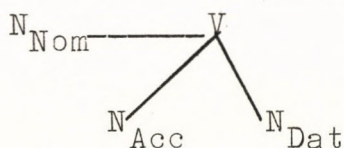
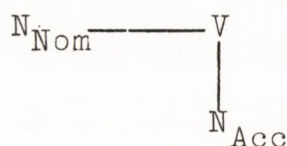
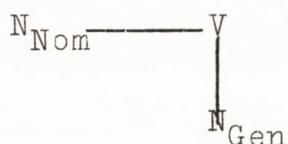
i.e. besides the subject an objective and a dative complement (a substantive or a nominal group) is to be expected.^{*}/



i.e. besides the subject an objective and an instrumental complement (a substantive or nominal group) is to be expected.^{*}/

^{*}/ The occurrence of the abovementioned complements is not necessary, only expected; we obtain, of course, correct sentences also when some of the complements fail to appear. E.g. "Пётр читает"; "Пётр читает книгу".

2.2 With verbs having different governments in Hungarian the coding contains, besides the code-number referring to the government, also a device referring to the corresponding "translation" into Hungarian. E.g.



2.3 The verbal complements expressed by a prepositional construction are not specific to the different verbs. The expression of the temporal, local etc. circumstances of any action or any event may occur in the sentence. In addition to adverbs, these circumstances may be expressed, by prepositional constructions being adverbial complements joined with the verb. A problem arises in connexion with the prepositional constructions when they stand behind the substantive or the substantival group, may also be the attributive complements of a substantive or a substantival group standing before it. In such cases further analysis is required. The decision of the attributive or adverbial role of the prepositional constructions standing behind the substantive or substantives demands a particular information-system based on the system of possible governments. In previous publication the present author dealt in some detail with a possible type of the coding of such constructions [7].

2.4 It is possible to decide whether the adverb in a given sentence is a complement of a verbal or adjectival morpheme without a special code-system only on the basis of special devices.

2.5 The complements of the substantives and adjectives having unvaried governments may be found out by the aid of particular code-numbers (управление + N_{Instr} равный + N_{Dat}; характерный + для + N_{Gen}).

2.6 The interconnection between the substantive in genitive and the substantive standing before it may be determined in the course of the morphological analysis. To decide whether a substantive in genitive (or a nominal group in genitive) represents really a possessive relation or only a quality (gen. qual.) is impossible without taking into account some semantic features, i.e. without semantic analysis, this problem goes, therefore, beyond the frame of syntactic analysis.

3. After the morphological and syntactical analysis roughly outlined here, it is obvious that there exist a great many cases when we cannot get the necessary information by means of a lower level analysis. In these cases the determination of the role of the different elements of the sentence is only possible if we also take into consideration the semantical features of these elements. That is what we call semantic analysis.

The most frequent cases in which semantic analysis is required are the following:

3.1 Substantives with endings of nominative and accusative (or nominal groups with such endings).

The nominative and the accusative of neuter substantives, of masculine inanimate substantives and of feminine substantives ending in ь are morphologically identical. In many cases the sentence turns out to have such a structure that the grammatical role of the substantive (nominal group) is self-evident without any analysis. So, for instance, if the verb of the sentence does not govern an object, the substantive (or the nominal group) stands doubtlessly in the nominative; ...явление происходит...

The same holds true, as a matter of course, if the given verb governs an object but, the verb stands in passive voice or is reflexive: ...деталь обрабатывается...

Further analysis is not needed also in such cases when between the two substantives (or nominal groups) belonging to the transitive verb only one turns out to be homographic:

...завод получил машину...
...машина вращает деталь...

In the abovementioned cases the determination whether the homograph substantive (or nominal group) stands in the nominative or accusative is only a question of programming.

From the point of view of analysis only those cases pose a problem in which both substantives (or nominal groups), belonging to the transitive verb, are homographs. Since the word order in Russian is not fixed to the same extent as, for instance, in English, on the basis of formal features any one of the two substantives may be considered as subject or object. E.g.

...основное уравнение динамики получает такой вид в том случае...

...такой вид получает основное уравнение динамики в том случае...

or:

...завод выпустил такой продукт...

...такой продукт выпустил завод...

Further investigations are required to decide whether it is possible to analyze mechanically (i.e. for the purposes of machine translation) such constructions by the aid of a suitable system of semantic information or whether another type of information is needed.

3.2 Possessive relations.

In cases of possessive relations the question to be answered is whether the substantive in genitive is a possessive adjective or a descriptive adjective of the substantive standing before it. (We shall leave aside the cases in which it is possible to translate the Russian possessive relation into Hungarian with a compound word or it is expressed in Hungarian by the aid of an adjectival ending, because in these cases the translation by a possessive relation does not result in ambiguities. E.g.:

давление пара; вопросы технологии.

In Russian the possessive relations expressing a descriptive adjective have two distinctive features:

- 1) The substantive in genitive has always an adjective;
- 2) the substantive in genitive concerns quantity, quality, measure etc. The quality expressed by the possessive word (the substantive in genitive) is connected with the possession: if we change the role of the two morphemes, we get a real possessive relation. E.g.:

сталь высокого качества; качество стали; пути одинаковой длины; длина пути.

Thus, the solution of our problem is as follows:

- 1) We have to analyze further those cases only, in which the possessive word has an adjective.
- 2) It has to be determined whether the possessive word denotes quality and whether this quality may be the quality of the possession.

The solution of these problems requires such a code-system on the basis of which it may be stated which substantives may form a real possessive relation and with which other substantives. The inversion of the role of the two substantives (possession \longrightarrow possessive word and vice versa) enables us to clear up the interconnection between the given possession and the quality expressed by the given possessive word. (But this is not an extra operation for the machine, it means only the comparison of the corresponding code-numbers.)

If the abovementioned conditions are not fulfilled (i.d. the possessive word has not an adjective; the possessive word does not denote quality; the possessive word does denote quality but a quality which cannot be attached to a possession), the construction must be translated as a possessive relation even if according to the code-system a real possessive relation is not possible. E.g. in the abovementioned "продукт завода" construction the substantives may occur also in an reversed order: "...завод этого продукта...". This construction does not express a descriptive adjectival relation (the word "продукт" does not denote quality), thus, although the code-system regards only the "продукт завода" possessive construction as suitable, nevertheless it must be translated by a possessive adjectival construction. If we should succeeded in compiling such a code-system which would refer also to these "reciprocal" possessive relations, this part of devices could be omitted.

The question could be solved on the basis of such a semantic information-system which would reveal the different qualities the given substantives can have. In possession of such a system, the algorithm for the possessive relations could be simplified to a great extent.

3.3 Verbs with the ending -ся

The translation of constructions containing as predicate a verbal morpheme with the ending -ся (or in a participial construction a participle with the ending ся) can only be executed, if the passive or reflexive role of the verbal morpheme is cleared up and if the morpheme fails to be an expression for impersonal subject. This is necessitated by the problem of interpretation of the verbal form and of the substantive in instrumental case (or of a corresponding nominal group) occurring in the construction. Namely, the substantive in instrumental designs the agents of the verb in a passive construction (since Hungarian has no passive constructions, this will be the subject of the Hungarian sentence), but it is an adverbial modifier of means or an adverb of manner in a reflexive construction or in a construction with an impersonal subject. E.g.:
...перевозка материалов обеспечивается применением машин...;
...механическая работа обеспечивается силовыми машинами....

There are among the abovementioned constructions a great many of which the passive or the reflexive character can be found out by suitable programming without any semantic analysis:

- 1) If the verb is a verb with perfective aspect and the construction is a reflexive one.
- 2) If the verb cannot be connected with an adverbial modifier of means (see point 2,1) and the substantive in instrumental in the construction is inanimate and the construction is a reflexive one. The substantive in instrumental case (or the corresponding nominal group) denotes in such cases an adverb of manner.
- 3) If the substantive in instrumental case in the given construction is animate and the construction is a passive one.

- 4) If the verb governs - contrary to Hungarian - the instrumental case; the establishment of the grammatical role of the substantive in the instrumental case (or of the corresponding nominal group) can be made on the basis of particular code-numbers (see point 2,2). Such verbs are: называться, являться, оставаться.
- 5) In the light of the above cases further analysis is only required if the verb in the construction is a verb with a progressive aspect and ends in -ся and the substantive in the instrumental case is inanimate.

The following considerations may render help to compile the semantic information required for further analysis:

- 1) Extra coding must be carried out for substantives in the instrumental case (or for the corresponding nominal groups) which are always used in the same form as adverbs of manner. E.g.: полностью, следующим образом, путём; ...подведённая работа используется полностью....
- 2) Extra code-numbers may be assigned to the verbs, of which the forms ending in -ся express an impersonal subject. E.g.: называться, выражаться, получаться.
- 3) If the substantive in the instrumental case of the construction denotes an abstract notion, we do not regard the construction as a passive construction. E.g.: ...перевозка материалов обеспечивается применением машин....
- 4) If the substantive in the instrumental case has a concrete meaning, the question must first be answered, whether the verb of the construction occurs among the constructions for expressing the impersonal subject. If the answer is affirmative, we may consider the construction as one with an impersonal subject and

translate the adverbial modifier of means into Hungarian with a substantive with the adverbial ending -val, -vel. E.g.: ...механическая работа выражается формулой...; ...работа...получается интегрированием отдельных элементарных работ... .

If the verb is not registered on the list of the verbs expressing an impersonal subject, we may consider the construction as a passive construction. E.g.: ...сопротивление развивается поднятым грузом...; ...механическая работа обеспечивается силовыми машинами... .

- 5) The solution is also feasible by comparison ^{*}/ of the two substantives of the construction (in accusative and instrumental case, respectively), taking into account, as a matter of course, the list of verbs expressing an impersonal subject. The basis of the comparison may be the possessive relation in the same way as we used it to clear up the grammatical role of the possessive relation (see point 3,2). If the substantive in the instrumental case may occur as the possessive word of the substantive in nominative, we consider the construction as a passive construction. Taking the above example we have: ...механическая работа обеспечивается силовыми машинами...

The possessive relation of the two substantives: работа машины. ...сопротивление развивается поднятым грузом...

The possessive relation: сопротивление груза.

The possessive relations such as "машина работы" and "груз сопротивления" do not occur among the possible ones

^{*}/ The expression "обеспечивать работу" must be coded extra as an idiom.

in our code-system, thus, according to the above we consider them as passive constructions in the translation.

If the possessive relation of the two substantives is only correct in reverse order (namely, the possessive word is the substantive in the nominative), the construction is a reflexive one or expresses an impersonal subject. E.g.: ...механическая работа выражается формулой...

The possessive relation: формула работы.

The verb occurs on the list of the verbs which express an impersonal subject, the sentence can be translated as follows: "A mechanikai munkát képlettel fejezzük ki". The translation is not incorrect even though we consider the construction - merely on the basis of the possessive relation of the substantives - as a passive construction. We receive: "A mechanikai munkát képlet fejezi ki". This kind of translation renders unnecessary the extra coding of the verbs which may express an impersonal subject. E.g.: ...перевозка материалов обеспечивается применением машин...

The possessive relation: применение перевозки.

The above example may also occur in the following manner: ...перевозка материалов обеспечивается машинами...

According to the above the construction may be regarded on the basis of the code-system, which allows the possibility for the two substantives of forming a possessive relation (перевозка машины), as a passive construction. The translation remains correct: "Az anyagok szállítását gépek végzik (biztosítják)".

It remains to be proved whether it is possible to determine the grammatical role of two substantives in a construction by means of the comparison based on the possessive relation in general. Up to the present we may only say that a great amount of examples shows the abovementioned regularities. From a pure-

ly practical point of view, it seems best, therefore, to use the above proposed solution all the more because the code-system and the devices applied for the solution of the problems in point 3,2 may also be used here in the same manner.

With respect to the translation of the verbal forms ending in -ся we have limited ourselves to such cases as are important from the point of view of semantic information. L.Dezső has recently tackled the question in detail (2,3). He also presumes in his algorithm semantic information.

3.4 Prepositional constructions.

The governments enable us to clear up whether the prepositional constructions standing after the substantive (or substantives) play an attributive or an adverbial role in the construction under consideration. This procedure is, however, not enough in itself: quite frequently semantical information is also required. E.g. the constructions ...установили печь на новом заводе... and ...установили печь на твёрдом топливе... are morphologically identical, it is impossible even on the basis of government to differentiate between them: the role of "на новом заводе" as an adverbial modifier of place and the attributive role of "на твёрдом топливе" can only be stated on the basis of the meaning of the verb and the substantive, respectively of the substantive standing before the prepositional construction and of the substantive in the prepositional construction: установить + на + a substantive denoting place: a construction with an adverbial modifier of place; печь + на + concrete noun: an attributive construction.

The solution of this problem requires a code-system which gives information on the lexical meaning of the given morpheme.

4. For the solution of the problems mentioned in point 3. the vocabulary of the machine has to be completed with code-numbers concerning the suitable semantic information. The code-system giving semantical information supposes the subdivision of the vocabulary according to the categories of meaning. The different morphemes belong to certain sets (groups) according to their lexical meaning. The semantic information is therefore such a code-number which indicates in which group the given morpheme belongs.

The possibilities of interconnections between the different groups must be a part of the program. The programs which serve to clear up the syntactic role of the morphemes must take into consideration in the syntactic analysis, beyond the mentioned possibilities of government, the semantical possibilities, too. The program containing the semantic information plays an important role, besides in the determination of the syntactic role of the morpheme, also in finding out the meaning of the given morpheme. (E.g.: ...reshít' zadachi za opredelénnoe remya... = "meghatározott időre"). Here the meaning of -ra, -re of the preposition "за" comes from the meaning of "time" of the substantive in the construction.

The possible possessive relations of the substantival morphemes may be compiled in a rather complicated table. This table, however, demands a great amount of capacity, it is, therefore, more suitable to build these constructions in to the program directly.

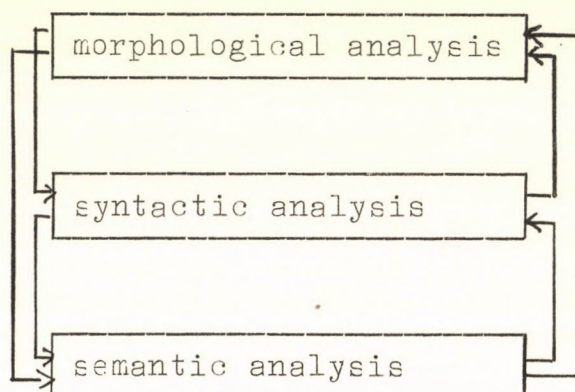
The number of groups necessary for the semantic information is determined by the size of the machine vocabulary, the length of the programs utilizing the semantic information is fixed by the requirements concerning the correctness of the translation. The more words the vocabulary contains and the more precise translation we can obtain, the more detailed the semantic information-system is and the longer program are required.

5. The code-system consists of the following parts:

- a) the code referring to the form class (C.I.)*)
- b) the code referring to morphological features (C.II.)*/,
- c) the code referring to the governments, necessary for the syntactic analysis (C.III.),
- d) the code referring to semantic information (C.IV.).

C.I., C.III. and C.IV. refer to permanent features of the morpheme, C.II. must be determined during the program runs [4,6].

The scheme of the mechanical analysis



The scheme shows the "structure" of the program required for the analysis. The program begins with the morphological analysis of the word forms of a given text. In the case of homographic forms also syntactic and semantic relations are taken into consideration. As a result C.II. will be stored.

The second step is the determination of the syntactical role of the different morphemes by the aid of C.III. In many cases governments fail to give the required information, thus

*/ The code-system C.I. and C.II. has been elaborated by Gy.Hell [4].

semantic information must also be taken into account (C.IV.). The result of the syntactic analysis or of the semantic one, if any, must be stored extra: this will be the structural formula of the Russian sentence under consideration, which enables the synthesis of the corresponding Hungarian sentence.

The proposed semantic information-system is by no means complete, it aims only at the clearing up of some syntactic constructions. A completely unambiguous and adequate Hungarian translation would require such a semantic information-system, on the basis of which it would be possible to state mechanically all the possible semantic interconnections of morphemes. This would make possible to build up not only the syntactic structure of the given sentence of the source language but also its semantic structure, i.e. his "semantic" formula. The whole analysis would deliver two ranks of information, a grammatical and a semantical one. The synthesis on the basis of their cross-checks would provide an unambiguous and quite correct sentence of the target-language.

6. Since both the compiling of the government-system necessary for the syntactic analysis and the word-sets as the base for the semantic analysis can be carried out only with respect to a given vocabulary, the practical work may begin only after the vocabulary has been compiled. But it can already be stated in advance that only some particular verbal and nominal morphemes subdivided according to the government, may belong to the same group, and the same holds true with respect to the subdivision according to the meaning of the morphemes for the purposes of semantic analysis. The comparison of these two groupings may lead to interesting results of great linguistic relevance. It would be interesting to inquire into the size of the different groups, to state with what kind of other morphemes a given morpheme will belong to the same group, to clear up the differences and coincidences between the two classes of

the given morpheme. From the result received in this way we should be able to draw some conclusions regarding the semantical content of the words. Thus, this comparison may give information about the interconnections between the lexical meaning of the given morpheme and its syntactic function [1].

REFERENCES

- [1] ANTAL, L.: A jelentés kérdései (The Questions of Meaning, in Hungarian), Nyr. 37, pp. 81-96.
- [2] DEZSŐ, L.: Машинный перевод русских конструкции с члаго-лами на -ся на венгерский язык. *Studia Slavica* 9, pp. 239-253.
- [3] DEZSŐ, L.: Az orosz szenvedő szerkezetek gépi fordításának nyelvtani kérdései (Questions Concerning the Translation of Russian Passive Constructions, in Hungarian), Nyr 87, pp. 446-453.
- [4] ХЕЛЛ, Г.: Определение номинальных групп в МП с русского на венгерский. *Computational Linguistics* I. pp. 5-106 (1963).
- [5] KÁROLY, S.: Tesnière szintaxisa és a szintaxis néhány kérdése (The Syntax of Tesnière and some Questions of Syntax, in Hungarian), *Ált. Nyelv. Tan. I.*, Akadémiai Kiadó, Budapest, 1963, pp. 161-186.
- [6] SIPŐCZY, GY.: Algorithms for Analysis of Homonymic Nominal Forms. *Computational Linguistics* I., pp. 109-187, (1963).
- [7] ШИПЕЦИ, Д.: Вопросы программирования русских предложных оборотов. *Studia Slavica Debrecensis* 3, pp. 59-72 (1963).
- [8] VARGA, D.: Morphological Analysis with the Help of the Method of Successive Delimitation, *Computational Linguistics* I., pp. 223-255 (1963).

ОПЫТ АВТОМАТИЧЕСКОЙ СЕГМЕНТАЦИИ ТЕКСТА ПРИ МП С РУССКОГО ЯЗЫКА НА ВЕНГЕРСКИЙ

И. Ботош

Условные обозначения

ε	сегмент
$\varepsilon 1$	сегмент первой степени
$\varepsilon 2$	сегмент второй степени
$\varepsilon 3$	сегмент третьей степени
$\varepsilon 1S$	именной сегмент первой степени
$\varepsilon 1V$	глагольный сегмент первой степени
$\varepsilon 1X$	сегмент первой степени другого типа
$\varepsilon 1S'$	$\varepsilon 1S$, падежная форма или предлог + падежная форма которого зависит от одного из предыдущих $\varepsilon 1S$. (Если не от непосредственно предыдущего, то зависимость указывается индексом в правом нижнем углу, напр.: $\varepsilon 1S + \varepsilon 1S^g + \varepsilon 1S'_{\varepsilon 1S}$ - третий компонент зависит от первого).
$\varepsilon 1S^p$	$\varepsilon 1S$, состоящий из A^p или $A_v + A^p$
$\varepsilon 1S^g$	именной сегмент первой степени без предлога, где $n/A/ + /S/$ имеет форму родительного падежа
$\varepsilon 2S$	именной сегмент второй степени
$\varepsilon 2V^{prAv}$	$\varepsilon 2V$ с одним лишь $\bullet V$, являющимся предикативным наречием
$\varepsilon 2V^{inf}$	$\varepsilon 2V$ с одним лишь V , являющимся инфинитивом

S	имя существительное, местоимение-существительное
Pr	предлог
A	согласуемое имя (полная форма прилагательных и причастий, согласуемые в роде, числе и падеже местоимения), все числительные, притяжательные местоимения "его", "ее", "их"
Av	наречие, частицы "более", "менее", "наиболее", "наименее", "не"
V	предикатив (<i>verbum finitum</i> , краткая форма прилагательных и причастий, предикативные наречия, предикативная форма местоимений), инфинитив, деепричастие
S ^g	слово категории S в форме родительного падежа
A ^g	слово категории A в форме родительного падежа
A ^p	причастие (полная форма)
C	запятая
Cj	сочинительный союз
ZI	левая скобка
ZII	правая скобка
Pa	частица (за исключением "не", "более", "менее", "наиболее", "наименее")
Sj	подчинительный союз
Pt	точка
Ve	вводное слово
T	тире
Pc	точка с запятой
-	(над знаком) отсутствие члена
I, II, III...	первое, второе, третье и т.д. слово, входящее в состав идиомы

/ граница s1
// граница s2
/// граница s3

Введение

При выработке программ МП немаловажную роль играет вопрос об автоматическом определении тех границ в тексте, в рамках которых имеют силу, отдельные указания программ анализа - и, в известной мере, и синтеза -, и за которыми они не имеют силу. Отрезки текста, помещающиеся внутри таких границ, принято называть сегментами. При выработке систем автоматического деления текста на сегменты исследователи, как правило, исходят из более крупных единиц - например, из фраз или фразообразных оборотов - и потом переходят на установление связей между элементами таких больших сегментов. Получается, таким образом, система сегментации "сверху вниз", от более крупных единиц к более мелким.*/

Имея в виду анализ текстов вообще, или даже анализ, примененный к целям МП с любого языка на любой, такой подход необходим, ибо разная определенность порядка слов в разных языках делает нужным предварительный анализ структуры целой фразы или фразообразного оборота, прежде чем перейти к получению информации, касающихся подробностей связи между более мелкими единицами текста. (Ср., например, условия порядка слов в английском, или, тем более, в немецком языках).

Исходя, однако, из конкретных задач МП с русского на венгерский и принимая во внимание относительно свободный

*/ Ср., например, доклад И.А. Мельчука на V Международном съезде славистов (I).

порядок членов предложения в обоих языках, мы можем попытаться построить систему с обратной последовательностью, т.е. систему определения сегментов "снизу вверх", от более мелких единиц к более крупным. При этом, наша попытка сегментации "снизу вверх" включает в себя элементы как метода непосредственных составляющих (immediate constituents), так и метода грамматики зависимостей ("хозяев" и "слуг", dependency grammar).

Естественной единицей для деления текста оказывается его отрезок от пробела до пробела. Такой отрезок будем условно называть "словом", подразумевая под этим термином и любой знак препинания (перед последним всегда мыслится пробел).

Анализ русского текста с точки зрения определения границ будет происходить на трех ступенях (определение сегментов первой, второй и третьей степени). Главной задачей анализа на первой ступени является определение связей между предложениями и словами, относящимися к разным категориям имен (именные сочетания), а также связей между предикативами и наречными словами (глагольные или предикативные сочетания). На основании такого анализа можно выбирать соответствующие указания для предварительного синтеза. При анализе на второй ступени делается упор на уточнение именных сочетаний, расширенных несогласованными определениями и причастными оборотами. С точки зрения синтеза текста на венгерском языке эта ступень имеет особенно большое значение, так как структура зависящих друг от друга именных сочетаний в обоих языках очень различна. Главной же целью анализа на третьей ступени является определение места сказуемого при синтезе. Упор здесь сделан на соотношения глагольных (или других предикативных) и следующих за ними именных сочетаний. Именные же сочетания, предшествующие глагольным, как правило, не имеют влияние на место сказуемого в венгерском языке, и поэтому нам кажется ненужным включать их в глагольные сегменты третьей степени.

сегмент первой степени (s1) состоит из одного или нескольких слов, сегмент второй степени (s2) - из одного или

нескольких сегментов первой степени, а сегмент третьей степени (ε_3) - из одного или нескольких сегментов второй степени.

В дальнейшем мы попытаемся установить те признаки сегментов, которые могут быть обнаружены путем автоматического анализа. Результаты данной попытки нельзя считать окончательными, они скорее представляют собой лишь первые шаги на пути установления надежных формальных признаков, тщательная проверка и дополнение которых будут еще необходимы. Текстовым материалом для настоящего исследования послужила статья Е.А. Жоголева "Система программирования с использованием библиотеки подпрограмм" (2). Примеры приводятся, главным образом, из этой статьи. Разумеется, что выводы, полученные на основании обработки одной лишь статьи (всего 38 страниц), нужно пока считать предварительными.

I. Сегменты первой степени

Сегментами первой степени (ε_1) мы назовем такие отрезки текста входного языка, автоматический анализ которых дает достаточное число информации для предварительного синтеза эквивалента на выходном языке. Анализом соотношений сегментов первой степени, входящих в состав сегмента второй степени, получается потом дополнительная информация, позволяющая уточнить некоторые подробности синтеза. Дальнейшее уточнение синтеза становится возможным на базе анализа сегментов третьей степени.

Для иллюстрации сказанного приводим следующий пример:

При вполне допустимых ограничениях
на класс
решаемых задач
оказалось
целесообразным

$$\left. \begin{array}{l} \varepsilon_1 \\ \varepsilon_1 \\ \varepsilon_1 \end{array} \right\} \varepsilon_2 = \varepsilon_3$$
$$\left. \begin{array}{l} \varepsilon_1 = \varepsilon_2 \\ \varepsilon_1 = \varepsilon_2 \end{array} \right\} \varepsilon_3$$

задавать
часть
информации
с помощью
внешних адресов

задание
информации
при каждом обращении
к подпрограмме
можно выполнять
различными способами

$$\left. \begin{array}{l} \varepsilon_1 = \varepsilon_2 \\ \left. \begin{array}{l} \varepsilon_1 \\ \varepsilon_1 \end{array} \right\} \varepsilon_2 \\ \left. \begin{array}{l} \varepsilon_1 \\ \varepsilon_1 \end{array} \right\} \varepsilon_2 \end{array} \right\} \varepsilon_3$$

$$\varepsilon_1 = \varepsilon_2 = \varepsilon_3$$

$$\left. \begin{array}{l} \varepsilon_1 \\ \varepsilon_1 \end{array} \right\} \varepsilon_2 = \varepsilon_3$$

$$\left. \begin{array}{l} \varepsilon_1 \\ \varepsilon_1 \end{array} \right\} \varepsilon_2 = \varepsilon_3$$

$$\left. \begin{array}{l} \varepsilon_1 = \varepsilon_2 \\ \varepsilon_1 = \varepsilon_2 \end{array} \right\} = \varepsilon_3$$

$$\varepsilon_1 = \varepsilon_2 = \varepsilon_3$$

Среди сегментов первой степени различаются:

- а) именные ($\varepsilon_1 S$)
- б) глагольные или предикативные ($\varepsilon_1 V$)
- в) других типов ($\varepsilon_1 X$)

Именные сегменты первой степени могут, например, состоять из:

- а) существительного (S) (составление),
- б) предлога и существительного (Pr + S) (для решения),
- в) согласуемого имени и существительного (A + S) (данной задачи),
- г) предлога, согласуемого имени и существительного (Pr + A + S) (между отдельными массивами).
- д) наречия, согласуемого имени и существительного (Av + A + S) (ранее составленные программы; весьма значительных результатов),*

* В работе Г.Хэлла (3), где вопрос обсуждается с точки зрения структуры, наречия не принимаются во внимание, так как они "с точки зрения структуры никакой роли не играют и поэтому не входят в структуру номинальных конструкций" (стр. 32). С точки зрения структуры это и понятно. В настоящей же работе, где на первый план выдвигается сегментационная сторона вопроса, определение принадлежности наречия к тому или иному сегменту приобретает весьма большое значение. Подобным образом, в формулах, данных в работе (3), порядок слов роли не играет (для структуры это не важно), а для сегментации он имеет важное значение.

- е) предлога, наречия, согласуемого имени и существительного (Pr + Av + A + S) (без ранее намеченной программы),
- ж) согласуемого имени (A) (другой - в контексте: "использование одной программы при составлении другой"),
- з) предлога и согласуемого имени (Pr + A) (в которых).

Под "согласуемым именем" (вернее, "категорией A") мы понимаем не только все склоняемые и согласуемые с именем существительным разряды слов (полные формы прилагательных и причастий; согласуемые в роде, числе и падеже местоимения), но и все числительные, а также употребляемые в качестве притяжательных местоимений формы "его", "ее", "их", которые на первой, предварительной стадии автоматического анализа различаются от омонимичных форм личных местоимений своим положением после предлога, слова категории A или существительного. Итак:

Pr + его (ее, их) = Pr + A (на его роль)
 A + его = A + A (без какой-либо их переработки)
 S + его = S + A (пути их преодоления)

В словарных статьях "его", "ее", "их" помещается информация, что в положении после Pr, A, S они относятся к категории A, а в других положениях - к категории S. Таким образом, в контексте "считать ее решенной" будут установлены три сегмента первой степени: slV (V - считать), $\text{slS}(S\text{-}ee)$, $\text{slS}(A\text{-} \text{решенной})$. Подобным образом, в контексте "если их рассматривать" слово "их" определяется как slS , состоящий из S.*

*/ В случаях типа "пути их преодоления" на основании словарного указания (пути + их : S + A) определяется разряд слова "их", а на основании правил первичной сегментации определяются сегменты первой степени (пути / их преодоления: $\text{slS}(=S)$ / $(\text{slS})(=A+S)$). Остается пока нерешенным вопрос об определении разряда слов "его", "ее", "их" перед S не после A или Pr. Ср., с одной стороны, "наметить их пути", а с другой: "считать его элементом структуры".

В разряд S, кроме имен существительных, входят также и местоимения-существительные (я, себя, кто, никто и т.п.).

Перечисленные до сих пор случаи можно объединить формулой:

$$Pr + Av + A + S,$$

где ни один из членов сегмента не является неизбежным; условием наличия $s1S$ считается только присутствие A или S и, при наличии нескольких членов, порядок их последовательности. Av входит в $s1S$ только в том случае, если за ним следует A.

Принимая во внимание случаи типа "данная перерабатывающая программа", "без какой-либо их предварительной ручной переработки", "все заранее заданные параметры", надо расширить формулу таким образом:

$$Pr + n (Av + A) + S,$$

где n обозначает возможность повторности сочетания Av + A (конечно, Av может в любом Цикле повторяемости равняться нулю, так же, как и A, ср. "очень хорошо известных").

Член сегмента может из себя представлять "идиому", т.е. состоять из нескольких слов (во время - $PrI + PrII$; тех или иных - $AI + AII + AIII$; тот же - $AI + AII$; одну и ту же - $AI + AII + AIII + AIV$). Соответствующие информации даются в словаре.

Несколько A могут быть разделены запятыми, а последние два также и сочинительным союзом. (Повторяющийся союз может находиться и перед первым A).

Получается формула:

$$Pr + Cj + Av + A + n (C + Cj + Av + A) + S,$$

где C = запятая, Cj = сочинительный союз.

Так, например, сегмент "внутренних и внешних адресов" соответствует формуле:

$$\bar{Pr} + \bar{C}j + \bar{A}v + \bar{A}_1 + \bar{C} + Cj + \bar{A}v + A_2 + S,$$

где черточка над обозначением члена имеет значение: "отсутствие члена".

Другой пример:

"и постоянные, и внутренние адреса"

$$\overline{Pr} + Cj_1 + \overline{Av} + A_1 + C + Cj_2 + \overline{Av} + A_2 + S.$$

Целесообразно с точки зрения определения сегментов считать слово "и" всегда союзом, даже в том случае, если оно в действительности является частицей, и различать разные функции "и" только при синтезе.

Av может обозначать не только наречие (в том числе и наречие меры: очень, весьма и т.п.), но также и частицу "не" и, кроме того, частицу, служащую для образования степеней сравнения (более, менее, наиболее, наименее). Разряд этих слов - Av; а для синтеза дается при них специальное указание. Итак, на-пример:

при составлении

другой, более сложной

$$\underbrace{Pr + S}_{\varepsilon lS}$$

$$\underbrace{A_1 + C + Av + A_2}_{\varepsilon lS}$$

Возможно, что Av будет встречаться не в соседстве слов, могущих быть членами εlS и не в соседстве глагола (в последнем случае оно войдет в состав εlV).

В таком случае данное Av будет автоматически определяться как $\varepsilon l\lambda$ (см. ниже).

В рамках εlS одно или несколько согласуемых имен может находиться в скобках, ср. в контексте: "при работе / другого (программного) массива". Для включения таких случаев в нашу формулу, следует ее видоизменить:

$$Pr + ZI + Cj + Av + A + ZII + n (C + ZI + Cj + Av + A + ZII) + S$$

(формула εlS)

где ZI символизирует левую скобку, а ZII-правую.

В нашем примере ["другого (программного) массива"]:

$$\overline{Pr} + \overline{ZI} + \overline{Cj} + \overline{Av} + A_1 + \overline{ZII} + \overline{C} + ZI + \overline{Cj} + \overline{Av} + A_2 + ZII + S$$

Итак, возможные члены ϵlS следующие: Pr , C , ZI , ZII , Cj , Av , A , S . Однако не все они равноправные. Есть среди них такие, которые и сами по себе могут составлять ϵlS и обязательно являются компонентами ϵlS ; это - A , S . Есть, которые сами по себе не составляют ϵlS и не обязательно являются компонентами ϵlS ; это - C , ZI , ZII , Cj , Av . И, наконец, есть разряд слов, сам по себе не составляющий ϵlS , но обязательно являющийся его компонентом, так как вне ϵlS он не встречается; это - Pr . Автоматическое обнаружение ϵlS происходит на основании обнаружения одного из его т.н. сильных компонентов (т.е. слов, обязательно входящих в состав ϵlS): Pr , A , S . Обнаружение же одного из слабых компонентов ϵlS (Av , Cj , C , ZI , ZII) не является признаком наличия ϵlS .

Первым компонентом ϵlS является:

- а) Pr (во всяком случае),
- б) ZI перед A или $Av + A$, не после Pr |
- в) Av перед A , не после Pr , ZI , Cj |
- г) A не после Pr , ZI , Av , Cj |,
- д) S не после Pr , ZI , A т.

Последним компонентом ϵlS является:

- а) A не перед S или $ZII + S$ |,
- б) ZII не перед S |,
- в) S (во всяком случае).

Итак, ϵlS , первым компонентом которого является S , состоит из одного лишь слова; ϵlS , первым компонентом которого является Pr , ZI , Cj или Av , состоит из нескольких слов; а ϵlS | первым компонентом которого является A , может состоять как из одного, так и из нескольких слов.

Глагольные (или предикативные) сегменты первой степени (ϵlV) могут состоять, например, из:

а) слова категории V; к этой категории принадлежат: *verbum finitum*, краткая форма прилагательного и причастия, предикативное наречие, предикативная форма местоимения, инфинитив, деепричастие (V) (упрощается; помещены),

б) двух слов категории V (V + V) (могут содержать; было осуществлено),

в) слова категории Av и слова категории V (Av + V) (не изменяются),

г) слова категории V и слова категории Av (V + Av) (функционирует нормально),

д) слова категории V, слова категории A и опять слова категории V (V + Av + V) (рассматривать не будем),

е) слова категории Av и двух слов категории V (Av + V + V) (не будем рассматривать),

ж) двух слов категории V и слова категории Av (V + V + Av) (можно осуществить автоматически),

з) трех слов категории V (V + V + V) (может быть составлена; должны быть расположены).

Объединяя все эти возможности сочетаний в одну формулу, получаем:

$$V + Av + V + Av + V$$

(формула $\varepsilon 1V$)

Итак, возможными компонентами $\varepsilon 1V$ являются: V, Av. Минимум $\varepsilon 1V$ — это V, способное и само по себе составить $\varepsilon 1V$ и обязательно входящее в состав $\varepsilon 1V$. Таким образом, V является сильным компонентом $\varepsilon 1V$, в то время как Av, неспособное само по себе образовать $\varepsilon 1V$ и не обязательно входящее в его состав, является его слабым компонентом.

Первым компонентом εlV является:

- а) V не после A_V ,
- б) A_V перед V , не после V .

Последним компонентом εlV является:

- а) V не перед A_V или V ,
- б) A_V не перед V или $A_V + V$.

Таким образом, εlV , начинающийся с A_V , обязательно состоит из нескольких слов, а εlV , начинающийся с V , может состоять как из одного, так и из нескольких слов.

Сопоставляя возможные компоненты именного и глагольного сегментов первой степени, мы находим среди них один общий возможный член: A_V . А если сопоставим список возможных первых и последних компонентов сегментов обоих типов, то увидим, что A_V может быть первым компонентом εlV и первым и последним компонентом εlV . Итак, на стыке двух сегментов первой степени могут возникать трудности насчет определения принадлежности A_V к тому или другому сегменту первой степени. Два εlV непосредственно друг за другом не следуют (в случае... $V + A_V + V...$ весь комплекс будет автоматически определяться как один εlV). На стыке $\varepsilon lS/\varepsilon lV$ также не возникает вопроса, поскольку - по формуле $\varepsilon lS - A_V$ на конце εlS не встречается. Вопрос о принадлежности A_V к тому или другому εl возникает на стыке $\varepsilon lV/\varepsilon lS$. Если εlS начинается с предлога, то предлог надежно указывает границу между двумя сегментами, являясь первым компонентом εlS . Например: "можно осуществить автоматически / с помощью...".

Но остается пока нерешенным вопрос о том, как определить принадлежность A_V к εlS или к εlV в таких случаях:

- 1) "использовать / ранее составленные программы"
- 2) "изучить основательно / приведенные данные"

Есть и такие случаи, когда даже человек затрудняется в определении принадлежности наречия, напр.: "закрепить навсегда определенный номер".

Сегменты первой степени других типов (εlX) состоят из:

а) Av , не входящего в состав εlV или в состав εlS (...и/тогда/адреса...),

б) Cj , не входящего в состав εlS (и),

в) Sj , т.е. подчинительного союза (если; для того чтобы),

г) частицы (Pa) (еще; только),

д) вводного слова ($V\varepsilon$), (вообще),

е) вводного слова и запятых ($C + V\varepsilon + C$), (,вообще говоря,)

ж) запятой (C), не входящей в состав εlS ,

з) одной из скобок (ZI ; ZII), не входящих в состав εlS ,

и) точки (Pt), тире (T), точки с запятой (Pc) и других знаков препинания.

Каждый εlX (за исключением типа $C + V\varepsilon + C$) состоит из одного лишь компонента, могущего со своей стороны состоять из одного или нескольких слов (ср. идиомы).

Итак, сильными компонентами εlX являются Sj , Pa , $V\varepsilon$ и знаки препинания, кроме запятой. Слабыми же компонентами εlX являются Av , Cj , C , ZI , ZII , не обязательно формирующие εlX , но могущие сами по себе составлять его (в том случае, если они не входят ни в один из других типов εl).

Приведем пример разделения текста на сегменты первой степени:

/Распределение / памяти / приходится делать / для каждой конкретной задачи / . / При этом/необходима / информация / как / о длинах / тех или иных массивов / , / так и / о характере / их использования / при решении / конкретных задач / . /

/ S / S / V + V / Pr + A + A + S / Pt / VsI + VsII /
V / S / Cj / Pr + S / AI + AII + AIII + S / C / CjI + CjII /
Pr + S / A + S / Pr + S / A + S / Pt /

/s1S / s1S / s1V / s1S / s1X / s1X / s1V / s1S /
s1X / s1S / s1S / s1X / s1X / s1S / s1S / s1S / s1S / s1X /

Другой пример:

/Вместе с информацией / о размещении / в памяти / массивов / , / участвующих / в решении / задачи / , / целесообразно вводить / и информацию / о значении / тех или иных параметров / программы / . /

/ Pr I + Pr II + S / Pr + S / Pr + S / S / C /
A / Pr + S / S / C / V + V / Cj + S / Pr + S / AI + AII + AIII +
+ S / S / Pt /

/s1S / s1S / s1S / s1S / s1X / s1S / s1S / s1S / s1X ,
s1V / s1S / s1S / s1S / s1S / s1X /

2. Сегменты второй степени

Именные сегменты второй степени (s2S) могут, например, состоять:

а) из s1S и одного или нескольких таких s1S, в которых нет предлога, запятой или союза и в которых n /A/ + S имеет форму родительного падежа.* / Формула такого s2S:

$$s1S + n /s1S^g/$$

* / В случае невозможности устранения морфологической омонимии, возможные формулы родительного падежа считаются формами родительного падежа.

В раскрытом виде:

$$\begin{aligned} & // Pr + Z I + Cj + Av + A + Z II + n (C + Z I + \\ & + Cj + Av + A + Z II) + S / + n [n (Z I + Av + A^g + Z II) + \\ & + S^g] // \end{aligned}$$

Примеры:

// начальное значение / данного параметра //

$$// \varepsilon lS / \varepsilon lS^g //$$
$$// A + S / A^g + S^g //$$

// после обработки / какого-либо массива //

$$// S lS / \varepsilon lS^g //$$
$$// Pr + S / A^g + S^g //$$

//использование / величин / какого-либо числового массива //

$$// \varepsilon lS / \varepsilon lS^g / \varepsilon lS^g //$$
$$// S / S^g / A_1^g + A_2^g + S^g //$$

// при работе / другого программного массива //

$$// \varepsilon lS / \varepsilon lS^g //$$
$$// Pr + S / A_1^g + Z I + A_2^g + Z II + S^g //$$

// для кодирования / числа / повторений / какой-либо групповой операции //

$$// \varepsilon lS / \varepsilon lS^g / \varepsilon lS^g / \varepsilon lS^g //$$
$$// Pr + S / S^g / S^g / A_1^g + A_2^g + S^g //$$

б) из εlS и одного или нескольких зависящих от главного его члена (от S ; при отсутствии S - от A) других εlS (не εlS^g). Зависимости этого рода определяются по информации в словаре об управлении имен (например, место в + предл. пад.; вопрос о и т.д. При причастиях - A^p - дается та же самая информация об управлении, что и при соответствующих глаголах). Формула такого $\varepsilon 2S$:

$$\varepsilon lS + n / \varepsilon lS' /$$

Примеры:

// переход / из одного состояния / в другое //

$$// \varepsilon lS / \varepsilon lS' / \varepsilon lS' \varepsilon lS //$$

// от его расположения / в памяти //

// $s_1 S$ / $s_1 S'$ //

в) из комбинаций случаев а) и б) :

$$(\alpha) s_1 S + s_1 S^g + s_1 S'_{s_1 S}$$

$$(\beta) s_1 S + s_1 S^g + s_1 S'_{s_1 S^g}$$

$$(\gamma) s_1 S + s_1 S' + s_1 S^g_{s_1 S}$$

$$(\delta) s_1 S + s_1 S' + s_1 S^g_{s_1 S'}$$

- где индексы в правом нижнем углу указывают на тот сегмент, от которого зависит данный сегмент.

Примеры:

// запись / результатов / в какой-либо числовой-массив //

(α)

// определение / его принадлежности / к последнему //

(β)

// с различным размещением / в памяти / таких массивов //

(γ)

// при обращении / к решению / конкретной задачи //

(δ)

г) из комбинаций случаев а), б) и в), например:

$$(\alpha) s_1 S + s_1 S^g_1 + s_1 S'_{s_1 S^g_1} + s_1 S^g_2 + s_1 S^g_{s_1 S^g_2} + s_1 S'_{s_1 S^g_2}$$

$$(\beta) s_1 S + s_1 S^g_1 + s_1 S^g_2 + s_1 S'_{s_1 S^g_1}$$

Примеры:

// изменение / размещения / в памяти / какого-либо / из этих массивов //

(α)

// при любом изменении / размещения / массивов / в памяти //

(β)

д) из $\varepsilon 1S$ и следующего за ним, зависящего от его главного члена (от S , при отсутствии S - от A) причастного оборота. Формула такого $\varepsilon 2S$:

$$\varepsilon 1S + \varepsilon 1X^C + \varepsilon 1S^P + n / \varepsilon 1S / + \varepsilon 1X^C$$

- где $\varepsilon 1X^C$ обозначает $\varepsilon 1X$, состоящий из запятой;

$\varepsilon 1S^P = \varepsilon 1S$, состоящий из $A_v + A^P$ (A_v , конечно, может равняться нулю);

A^P = полная форма причастия.

Если в процессе анализа встречается точка, то правая граница $\varepsilon 2S$ определяется перед ней.

Пример:

// программные массивы / , / не удовлетворяющие /
этим требованиям / , //
// $\varepsilon 1S$ / $\varepsilon 1X^C$ / $\varepsilon 1S^P$ / $\varepsilon 1S'$ / $\varepsilon 1X^C$ //

е) Разновидностью типа д) является:

$$\varepsilon 1S + \varepsilon 1X^C + \varepsilon 1S^P + \varepsilon 1S' + n / \varepsilon 1S^g / + \varepsilon 1X^C$$

Напр.: // массив / , / используемый / при решении / данной задачи / , //

// любой код / , / записанный / в соответствующих разрядах / команды / , //

// адрес / , / используемый / для кодирования / числа / повторений / какой-либо групповой операции //.

ж) Другой разновидностью типа д) является:

$$\varepsilon 1S + \varepsilon 1X^C + \varepsilon 1S^P + \varepsilon 1S' + n / \varepsilon 1S'' / + \varepsilon 1X^C$$

- где $n / \varepsilon 1S'' /$ указывает на сегменты, зависящие от дополнения причастия или друг от друга.

Суммируя все эти типы и, кроме того, имея в виду, что причастный оборот (без запятой) может находиться и перед εlS , к которому он относится, получим формулу $\varepsilon 2S$:

$$\begin{aligned} & \varepsilon lS^p + \varepsilon lS' + n / \varepsilon lS^g / + n / \varepsilon lS' / + n / \varepsilon lS^g / + \\ & + \varepsilon lS + n / \varepsilon lS^g / + n / \varepsilon lS' / + n / \varepsilon lS^g / + \varepsilon lX^c + \\ & + \varepsilon lS^p + \varepsilon lS' + n / \varepsilon lS^g / + n / \varepsilon lS' / + n / \varepsilon lS^g / + \\ & + \varepsilon lX^c \end{aligned}$$

Примеры автоматической сегментации по этой формуле (определение $\varepsilon 2S$):

// при фиксированном размещении / в памяти /
этих массивов /, / связанных / каким-либо образом / с первыми / , //

$$\begin{aligned} & // \varepsilon lS / \varepsilon lS' / \varepsilon lS^g_{\varepsilon lS} / \varepsilon lX^c / \varepsilon lS^p / \varepsilon lS' / \\ & \varepsilon lS'_{\varepsilon lS^p} / \varepsilon lX^c // \end{aligned}$$

что соответствует общей формуле:

$$\begin{aligned} & \overline{\varepsilon lS^p} + \overline{\varepsilon lS'} + \overline{\varepsilon lS^g} + \overline{\varepsilon lS'} + \overline{\varepsilon lS^g} + \overline{\varepsilon lS} + \overline{\varepsilon lS^g} + \\ & + \overline{\varepsilon lS'} + \overline{\varepsilon lS^g} + \overline{\varepsilon lX^c} + \overline{\varepsilon lS^p} + \overline{\varepsilon lS'} + \overline{\varepsilon lS^g} + \overline{\varepsilon lS'} + \\ & + \overline{\varepsilon lS^g} + \overline{\varepsilon lX^c}. \end{aligned}$$

// случай / зависимости / кода / операции / от
размещения / массивов / в памяти //

$$\begin{aligned} & // \varepsilon lS / \varepsilon lS^g_1 / \varepsilon lS^g_2 / \varepsilon lS^g_3 / \varepsilon lS^p_{\varepsilon lS^g_1} / \varepsilon lS^g_4 / \\ & \varepsilon lS'^2_{\varepsilon lS'_1} // \end{aligned}$$

По формуле:

$$\begin{aligned} & \overline{\varepsilon lS^p} + \overline{\varepsilon lS'} + \overline{\varepsilon lS^g} + \overline{\varepsilon lS'} + \overline{\varepsilon lS^g} + \overline{\varepsilon lS} + \overline{\varepsilon lS^g} + \\ & + \overline{\varepsilon lS^g} + \overline{\varepsilon lS^g} + \overline{\varepsilon lS'} + \overline{\varepsilon lS^g} + \overline{\varepsilon lX^c} + \overline{\varepsilon lS^p} + \overline{\varepsilon lS'} + \\ & + \overline{\varepsilon lS^g} + \overline{\varepsilon lS'} + \overline{\varepsilon lS^g} + \overline{\varepsilon lX^c} \end{aligned}$$

Первым компонентом $\varepsilon 2S$ является $\varepsilon 1S$ (любого типа).

Последним компонентом $\varepsilon 2S$ является:

- а) $\varepsilon 1S$ (любого типа) не перед $\varepsilon 1X^C + \varepsilon 1S^P$
- б) $\varepsilon 1X^C$ в случае, если предшествует $\varepsilon 1X^C + n / \varepsilon 1S /$

Возможными компонентами $\varepsilon 2S$ являются $\varepsilon 1S$ (любого типа) и $\varepsilon 1X^C$. Два соседних $\varepsilon 1S$ могут не входить в один $\varepsilon 2S$, если их соотношение не соответствует формуле $\varepsilon 2S$. Но каждый $\varepsilon 1S$ обязательно входит в состав какого-нибудь $\varepsilon 2S$ и способен и сам по себе составить $\varepsilon 2S$, являясь, таким образом, его сильным компонентом. Что же касается $\varepsilon 1X^C$, он является слабым компонентом $\varepsilon 2S$, входящим в его состав только при определенном условии - перед $\varepsilon 1S^P$ и после $\varepsilon 1S^P + n / \varepsilon 1S /$.

Глагольные сегменты и сегменты третьего типа второй степени тождественны с соответствующими сегментами первой степени.

$$\varepsilon 2V = \varepsilon 1V$$

$$\varepsilon 2V = \varepsilon 1X, \quad \text{не входящий в состав } \varepsilon 2S$$

3. Сегменты третьей степени

При выработке правил сегментации третьей степени главной целью было получение достаточных информации для определения места сказуемого в венгерском переводе. Этой цели соответствуют следующие правила:

1. $\varepsilon 3S = \varepsilon 2S$, не входящий в состав $\varepsilon 3V$

2. $\varepsilon 3V =$ а) если состав $\varepsilon 2V$ полон (т.е. или в нем нет слова, требующего инфинитива, или находится в нем и инфинитив):

$$\varepsilon 2V + n / \varepsilon 2X / + n / \varepsilon 2S / + n / \varepsilon 2X / + n / \varepsilon 2S /$$

б) если состав одного из $\varepsilon 2V$ не полон (не хватает инфинитива или предикативного наречия):

$$(\alpha) \quad \varepsilon 2V^{prAv} + n / \varepsilon 2X / + n / \varepsilon 2S / + \varepsilon 2V^{inf} + n / \varepsilon 2X / + n / \varepsilon 2S /$$

$$(\beta) \quad \varepsilon 2V^{inf} + n / \varepsilon 2X / + n / \varepsilon 2S / + \varepsilon 2V^{prAv} + n / \varepsilon 2X / + n / \varepsilon 2S /,$$

где $\varepsilon 2V^{prAv}$ обозначает $\varepsilon 2V$ с одним лишь V , являющимся предикативным наречием; а $\varepsilon 2V^{inf}$ обозначает $\varepsilon 2V$ с одним лишь V , являющимся инфинитивом.

Перед Pt или $C + Sj$ всегда определяется правая граница $\varepsilon 3V$.

3. $\varepsilon 3X = \varepsilon 2X$, не входящий в состав $\varepsilon 3V$.

Примеры:

/// Этот же оператор /// используется // после не-
которого его изменения // для определения / очередного свобод-
ного массива / при размещении / пассивных массивов /// . ///

/// $\varepsilon 2S$ /// $\varepsilon 2V$ // $\varepsilon 2S$ // $\varepsilon 2S$ /// $\varepsilon 2X$ ///

/// $\varepsilon 3S$ /// $\varepsilon 3V$ /// $\varepsilon 3X$ ///

/// В дальнейшем /// будем рассматривать // толь-
ко // стандартные массивы /// , /// хотя /// иногда можно //
вместе со стандартными // использовать // также и определенные
массивы /// . ///

/// $\varepsilon 2S$ /// $\varepsilon 2V$ // $\varepsilon 2X$ // $\varepsilon 2S$ /// $\varepsilon 2X$ /// $\varepsilon 2X$ ///

$\varepsilon 2V^{prAv}$ // $\varepsilon 2S$ // $\varepsilon 2V^{inf}$ // $\varepsilon 2S$ /// $\varepsilon 2X$ ///

/// $\varepsilon 3S$ /// $\varepsilon 3V$ /// $\varepsilon 3X$ /// $\varepsilon 3X$ /// $\varepsilon 3V$ /// $\varepsilon 3X$ ///

x

Пользуясь методом сегментации "снизу вверх" при МП с русского на венгерский, мы можем:

а) на первой ступени анализа получить информации, способствующие синтезу мелких сегментов, учитывая особенности элементарных именных и глагольных (предикативных) сочетаний;

б) на второй ступени анализа получить информации для синтеза всех именных конструкций, включая и самые сложные; с точки зрения синтеза на венгерском языке в этом отношении особенно важны разные типы именного управления (несогласованные определения, выраженные формой родительного падежа; "цепи" таких определений; специальные управления отдельных имен, в соответствии и вне соответствия с управлением глаголов того же корня) и причастные обороты, внутри которых глагольное управление причастий обрабатывается как именное управление, совпадающее с глагольным (тем же механизмом, что и любое именное управление - $s1S + s1S'$);

в) на третьей ступени получить информации относительно места сказуемого в предложении на выходном языке, а также уточнить ряд информации, полученных на более низких ступенях анализа.

ЛИТЕРАТУРА

- [1] МЕЛЬЧУК, И.А.: Автоматический анализ текстов (на материале русского языка). Славянское языкознание. Доклады советской делегации. Изд. АН СССР. Москва, 1963. (Отдельный оттиск).
- [2] В сборнике статей "Система автоматизации программирования", под ред. Н.П. Трифонова и М.Р. Шура-Бура. (Библиотека прикладного анализа и вычислительной математики). Гос. изд. физико-матем. лит. Москва, 1961.
- [3] ХЭЛЛ, Г. Определение номинальных групп в МП с русского на венгерский. Computational linguistics. I. Budapest, 1963.

A P P E N D I X

MATHEMATICAL LINGUISTICS AND MACHINE TRANSLATION IN THE COMPUTING CENTRE OF THE HUNGARIAN ACADEMY OF SCIENCES

by F. Kiefer

Historical

The work on MT from Russian into Hungarian began in 1961 when two teachers of Russian, Gy.Hell and Gy.Sipőczy, made an experiment at the University of Technical Sciences. This experiment did not go, however, beyond a word-by-word translation.

In March 1962 a conference was held jointly by the Computing Centre of the Hungarian Academy of Sciences, the Institute for Linguistics of the Hungarian Academy of Sciences and the Committee for Cybernetics of the Presidency of the Hungarian Academy of Sciences. This conference heard Hungarian research-workers in the field of mathematical linguistics and machine translation give account of their results and discussed some general problems. Gy.Hell and Gy.Sipőczy outlined their algorithms, S.Kónyi pointed out some mistakes in Mel'chuk's algorithm for MT from Hungarian into Russian, Gy.Szépe discussed in a paper some general aspects of MT. Besides the questions of MT, problems of mathematical linguistics were also discussed. Some papers dealt with algebraic linguistics (L.Kalmár, F.Papp and F.Kiefer), some with statistical linguistics (Including information theory; A.Rényi, I.Fónagy and others). Naturally, statistical linguistics has an old tradition in Hungary, in the field of algebraic linguistics, however, - as it was documented at the conference - a great gap has to be filled.

In the summer 1962 the MT-group was formed at the Computing Centre within the Department for Theoretical Questions. The group consisted of one full-time, one part-time mathema-

tician and of one part-time linguist. The first task of the group was the programming and checking of the algorithms elaborated by Hell and Sipőczy. The Computing Centre has a computer M-3 which is suitable for experimental purposes but does not make the checking of longer programs possible. In the case of long programs the program must be divided into several parts and checked successively. The other practical difficulty concerns the programming: the writing of programs demands more time than the elaboration of algorithms, in the first place because we have no suitable auto-code-system at our disposal and to write programs in machine language is a time-consuming affair. These circumstances determine to some extent the further research-work of the MT-group, giving it a more theoretical direction.

In September 1962 the Bolyai Society organized a colloquium on the topics "The Foundations of Mathematics, Mathematical Machines and their Applications". A separate session was devoted to mathematical linguistics. On the conference also Y.Bar-Hillel (Israel), R.B.Lees (USA), M.Bierwisch (DDR), S.Abraham (Roumania) and P.Sgall (Czechoslovakia) took part and gave lectures on different aspects of mathematical linguistics. As far as Hungarian research-work is concerned, Gy.Hell, L.Dezső and Gy.Sipőczy gave account of their results with respect to the mechanical analysis of Russian texts, F.Kiefer compared Kulagina's set theory model with Chomsky's generative model.

Current research

In September 1963 the MT-group was created a separate unit with an independent research plan with F.Kiefer as the leader of the group. At present the group consists of four full-time mathematicians (plus one guest-mathematician who is spending a year in our Institute) or rather "mathematical-linguists" because three of the mathematicians have also

linguistic qualification) and of two part-time linguists. Besides, a few outside linguists also take part in the work of the group.

The group co-operates with the linguistic group at the University of Debrecen (see below) and with the group for general and structural linguistics of the Institute for Linguistics of the Hungarian Academy of Sciences.

The current research covers the following fields of study.

a.) Machine Translation

First, the algorithms elaborated by Hell and Sipőczy were tested on the computer M-3 and later some parts of them on the computer Elliott-803, too. With respect to some special problems additional algorithms have been elaborated by B.Dömölki and F.Kiefer (dictionary look-up based on a method developed by Cooper, cutting of endings etc.).

All the elaborated algorithms up to the present do not essentially go beyond morphological analysis. The same holds true of the algorithm prepared by D.Varga.

The algorithms worked out by our group so far have all the following faults:

- i.) A recurring problem is that of storage-capacity, which is not a linguistic problem;
- ii.) Each algorithm is influenced by the computer available to us;
- iii.) Some of them are based almost exclusively on the binary system;

- iv.) The algorithms are worked out with high redundancy in that they do not contain essentially new aspects or answers besides solving some purely technical matters;
- v.) The morphological analysis is of less interest at the present state of MT.

Quite naturally then our group has recently turned to the questions of syntactic analysis. At the same time the investigations of semantic problems have come to the fore. For resolving the problems of syntactic analysis partly some purely empirical, partly some methods of model construction are used.

F.Kiefer and S.Abraham are seeking for a general model for MT on the basis of transformational grammar. The Comit-system is used for programming. The first results showed that the translation from English into Rumanian - on the basis of a very restricted corpus, of course - is far easier than the translation from English into Hungarian because of the approximate adequateness and inadequateness of the set of transformations.

A particular group headed by D.Varga has recently tackled some problems of syntactic analysis of Russian on the basis of technical texts. The texts are examined according to different aspects, e.g. agreement-problems, government, word order, participial constructions, verbal constructions etc. The results obtained in this way are important not only from the point of view of MT but also from that of language teaching and the theory of translation.

b.) Algebraic linguistics

F.Kiefer tackled Kulagina's set theoretic model in his first papers and applied it to Hungarian. He found out also the relations between Kulagina's model and Chomsky's transformational grammar and Bar-Hillel's quasi-arithmetical notation, respectively. Some relevant aspects of model construction has re-

gently been pointed out. F.Kiefer and S.Abraham work on a semantic model based on generative grammar. The model assigns to each word a matrix containing semantic categories and also some semantic categories for kernel sentences have been stated.

S.Abraham is also dealing with some of the interconnections of the theory of grammar and the theory of algorithms.

It is, however, still too early to correctly estimate the importance of this work and of the results obtained so far.

c.) Statistical linguistics

In the field of statistical linguistics - at least for the time being - there is no independent research work. As our group co-operates also with the Institute for Linguistics of the Hungarian Academy of Sciences we write also programs for statistical investigations but the interpretation of the results obtained does not come within our scope.

d.) The teaching of mathematical linguistics

F.Kiefer lectures at the Philological Faculty of the University of Budapest for mathematical linguistics for the fourth semester. Mostly students of mathematics and philology attend his lectures as well as teachers and research-workers. The number of attendants is between 15 and 20.

The lecture course has included the following subjects: general aspects of mathematical linguistics, model construction in linguistics, the set theoretic model of Kulagina, the model of Ajdukiewich-Bar-Hillel-Lambek, generative grammar, some further models (predictive analysis and push-down store grammars, projective grammars, dependency grammars etc.), statistical linguistics (above all, the ideas of Herdan and Mandelbrot). Some dissertations are expected in these topics in the near future. From the next semester a particular branch will be called into being at the University for training "mathematical-linguists".

e.) Further works

The MT-group publishes a periodical called "Computational Linguistics" twice a year (chief editor: F.Kiefer, editorial board: F.Papp, S.Petőfi, Gy.Szépe and D.Varga). The periodical gives account of the results of the group and of others being connected with our work. From this issue on foreign research-worker's papers will also be printed in this periodical.

In cooperation with Skriptor (Stockholm), bibliographical service is done at the Centre, i.e. bibliographical cards including all works of mathematical linguistics are edited and sent out to the subscribers of SMIL (Statistical Methods in Linguistics) and to others being on an exchange scheme of publications with our Centre.

F.Kiefer has been appointed Hungarian editor of SMIL which, from the next issue onwards, will deal not only with statistical methods in linguistics but with other methods also e.g. algebraic linguistics etc.

REFERENCES

- KIEFER, F., 1) About a set theoretic model in linguistics, (in Hungarian), in Ált.Nyelv.Tan.I. pp.187-200, Akadémiai Kiadó, Budapest (1963)
- 2) The Questions of Mathematical Linguistics (in Hungarian), in Valóság 6/1, pp.84-90 (1963)
- 3) Algorithmus und Programm für das Aufsuchen eines Wortes im Speicher, in CL 1, pp.189-201 (1963)
- 4) Algorithmus für die Analyse von Endungen, in CL 1, pp.203-221 (1963)
- 5) Bemerkungen zur Anwendung mengentheoretischer Methoden in der Sprachwissenschaft, in The Foundations of Mathematics, Mathematical Machines and Their Applications (L.Kalmár, ed.), forthcoming
- 6) Set-theoretic and Mathematico-logical Models in Linguistics (in Hungarian), in Ált.Nyelv.Tan.II., pp. , Akadémiai Kiadó Budapest, forthcoming
- 7) Some Aspects of Model Construction in Linguistics, in SMIL (Stockholm), forthcoming
- 8) On categorial grammars, (in Hungarian) in Ált.Nyelv.Tan.III., pp. , Akadémiai Kiadó Budapest, forthcoming
- VARGA, D., 1) Morphological Analysis by Help of the Method of Successive Delimitation in CL 1, pp. 223-255 (1963)

- 2) Проблемы осуществлению морфологического анализа в МП, in Научно - техническая информация, ВИНТИ (forthcoming)

AGÓCS, I., VARGA, D., Взаимосвязь элементов отношения русских и венгерских имён, in CL 2, pp. (1964)

- HELL, GY.,
- 1) Some Theoretic and Practical Aspects of MT from Russian into Hungarian; (in Hungarian), in Ált.Nyelv.Tan.II., (forthcoming)
 - 2) О некоторых характерных чертах алгоритма машинного перевода с русского языка на венгерский, in Slavica II (Debrecen), pp. 151-164 (1962)
 - 3) Определение номинальных групп в МП с русского на венгерский, in CL 1, pp. 5-107, (1963)
 - 4) The two Levels of the Syntactic Analysis in the Mechanical Analysis of Russian Texts, in CL 2, pp. (1964)

- SIPŐCZY, GY.,
- 1) The Programming of Prepositional Constructions (in Hungarian), in Ált.Nyelvt.Tan.II., (forthcoming)
 - 2) Algorithms for the Analysis of Homonymic Nominal Forms, in CL 1, pp. 109-187, (1963)
 - 3) Вопросы программированию русских предложных оборотов, in Slavica III. (Debrecen), pp. 151-169, (1963)
 - 4) Some Semantic Aspects of the MT from Hungarian into Russian, in CL 2, pp. (1964)

BOTOS, J., Опыт автоматической сегментации текста при МП
с русского языка на венгерский, in CL 2, pp.
(1964)

PETŐFI, S. J., Graphs in Linguistics, (in Hungarian), in Ált.
Nyelv. Tan. III., (forthcoming)

MATHEMATICAL AND APPLIED LINGUISTICS AT THE UNIVERSITY
OF DEBRECEN

by F.Papp

At the Department of Russian assistant professor Ferenc Papp, began with a seminar class entitled "Russian of to-day" from the academic year 1958. The students attend these classes for one (or two years). They get acquainted with the methods of modern descriptive linguistics through reading of recommended works and the accounts given by their fellow-students, and they do some research work themselves, too. The seminar activity consisted in statistical investigations of language in the first years (cf. the publications of F.Papp below); later a group made experiments with of transformational grammar on Russian sentences for about a half year or so; several groups worked on the compilation of the strong governments of the Russian and Hungarian verbs, respectively; one group examined Yngve's depth hypothesis with respect to Hungarian and Russian. From 1959 a number of theses have been elaborated on the topics dealt during the seminars.

Later on some teachers also joined the group consisting up to that time of only one teacher and his students. The collective worked as an independent research-unit under the name Work-group for Mathematical and Applied Linguistics of Debrecen from the beginning of 1963. The group is headed by Ferenc Papp. Similarly at the beginning of 1963 an agreement concerning the co-operation was brought about between the group of Debrecen and the Computing Centre of the Hungarian Academy of Sciences. According to the agreement the group of Debrecen conducts research work in MT and the Computing Centre renders help in programming.

At present the Debrecen group is engaged on the following research projects:

1.) Basic research in to the structure of Russian (and partly of Hungarian), with a theoretical aim, on the one hand and with keeping in view the possible applications for MT (statistical linguistics, transformational analysis, Ferenc Papp), on the other.

2.) The documentation including mathematical linguistics and MT (this work will be done jointly with the Computing Centre, the Debrecen group specializing in Russian material, F.Papp, S.Jánoska, J.Klauszer).

3.) The elaboration of algorithms for MT from Russian into Hungarian (jointly with the Computing Centre J.Klauszer).

4.) Modern approaches to language teaching and the theory of translation (F.Hadas, F.Papp).

5.) The compilation of an a tergo dictionary of Hungarian (together with the Computing Centre F.Papp and Á.Sebestyén).

The Debrecen group has the possibility to write articles for the journal 'Slavica' published by the Slavic Philological Institute of the University of Debrecen, hence most of the papers written by the members of the group appeared or will appear in this journal.

BIBLIOGRAPHY

We arrange the bibliographical material according to points 1 to 6 above. The themes 5.) and 6.) are only under preparation, no publications have appeared yet.

Ferenc Papp:

- Mathematical Methods in Linguistics (in Hungarian), -
A MTA Nyelv- és Irodalomtudományi Osztályának Közlemé-
nyei. XVII (1961), 1-4:289-306.
- Concerning Stylostatistical Method (in Hungarian), -
Filológiai Közlöny 1961.1-2:69-85.
- Трансформационный анализ русских присубстантивных конструкций
с зависимой частью - существительным. - Slavica
I(1961) 55-83.
- Количественный анализ словарной структуры некоторых русских
текстов. - Вопросы языкознания X(1961), 6:93-100.
- Некоторые вопросы изучения устной и письменной разновидностей
языка. - Slavica III(1963), 21-30.
- The Transformational Analysis of the Postpositional Construc-
tions with "után" (in Hungarian) -
Általános nyelvészeti tanulmányok I. Szerk.: Telegdi
Zsigmond. - (1963) - 237-61.
- К построению одной структурной модели системы словосочетаний
современного русского языка. - Studia Slavica IX.
(1963), 1-4:229-37.
- Как обнаруживается динамика языка при синхронном анализе? -
Славянска филология т. I, София 1963. - 222-3.
- Каково соотношение статистических методов с другими методами
изучения славянских методов? - Славянска филология,
210-1.
- Каковы основные задачи статистического анализа славянских
языков? - Славянска филология, 198-9.

Каковы различия между отдельными лингвистическими традициями в понимании грамматических явлений и каковы возможности унификации этих различных точек зрения? - Славянская филология, 245-6.

2.

(The bibliographical service of the Computing Centre - in form of bibliographical cards - includes also the work done in Debrecen; some other works are reviews only the more general papers are mentioned here:)

Ferenc Papp:

New Trends in Soviet Linguistics - Nyelvtudományi Közlemények LXL(1958), 2:392-410.

Judit Klauszer:

Вопросы математической лингвистики и МП на вежеских конференциях в 1962 гг. - Slavica III(1963), 149-50.

Ferenc Papp - Judit Klauszer:

Mathematical and Applied Linguistics in the Soviet-Union. - Nyelvtudományi Közlemények LXV(1963), 2:456-63.

Ferenc Papp:

Mathematical Linguistics and Machine Translation in the Soviet-Union - 0000 o. - Budapest 1964. (Kézirat)

3.

Judit Klauszer:

Some Problems of Synthesis of Hungarian Noun Forms -
Computational Linguistics II.

4.

Ferenc Papp - József Drahos:

About the Use of the Russian Pronoun свой (in Hungarian) -
Acta Univ. Debrecen 1961, 143-8.

Ferenc Papp:

Некоторые вопросы обучения пониманию русской речи на слух. -
Русский язык в национальной школе (Москва) 1963,
2:66-70.

Классификация русских глаголов. - Русский язык в школе (Москва)
24(1963), 4:98-101.

Ferenc Hadas:

Russian-Hungarian Grammatical Dictionary (in Hungarian) -
Felsőoktatási Jegyzetellátó Vállalat (1961) 164 l.

The Russian School Vocabulary (in Hungarian) - Pedagógiai
Szemle (1961) 6:501-512.

The Russian Dictionary for Schools (in Hungarian) - Az Idegen
Nyelvek Tanítása, (1961) 3:88-92.

Some Statistical Data and their Use in Teaching (in Hungarian) -
Az Idegen Nyelvek Tanítása (1962) 4:117-124.

О русско-венгерском грамматическом словаре, Slavica II (1962)
165-176.

Анализ уровня знаний студентов по русскому языку на первом
курсе в университете, Slavica III (1963) 192-204.

Э.А.Штейнфельдт:

Словарь-минимум русского языка для II-IV классов эстонских
школ, Таллин, 1961, 85 стр. Slavica III (1963) 170-173.

Pedagogy and Cybernetics (in Hungarian) - Valóság, (1963) 6.

MACHINE TRANSLATION WORK BY THE COMPUTER ELLIOTT 803

by M. Stein

I wish here to give an account of the solution of some problems connected with MT by the computer ELLIOTT 803.

The computer ELLIOTT 803 has the advantage over other available machines, that it has, on the one hand, a memory with larger capacity, and a machine word containing 39 bits, on the other.

Transliteration

The words are punched on tapes. Before punching the Russian text must be transliterated, i.e. the Cyrillic alphabet must be transcribed into letters used by Siemens. We use the following transliteration:

а	а	л	l	ч	4
б	b	м	m	ш	s
в	v	н	n	щ	w
г	g	о	o	ъ	5
д	d	п	p	ы	y
е	e	р	r	ь	6
ё	e	с	2	э	3
х	x	т	t	ю	q
з	z	у	u	я	9
и	i	ф	f		
и	j	х	h		
к	k	ц	c		

Input

Six bits come up for each of the signs in the words of the machine. The sixth sign (from right to left) is one or zero, accordingly to whether the sign in question is a letter or a figure. So a bit-combination of six signs can be put in a machine word. The remaining three bits may be used to store different data.

As the longest word may consist of 18 letters at most, three cells are required (maximum) to store one word. In the cell the 37th bit is one, if the content of the following cell is a continuation of the preceding one.

The words with different length are placed in 18 different groups: B1, B2, ..., B18. These are the initial addresses of cells, which the words consisting of 1, 2, ..., 18 letters, respectively, are put in.

The place in the storage of the given word can be computed,

$$B_j + \alpha I_j \quad (1 \leq j \leq 18)$$

where $\alpha = 1, 2$ or 3 accordingly to whether the word in question needs 1, 2 or 3 cells. It shows how many words, consisting of 1, 2, ..., 18 letters, have already been put in.

Storage

The words of the different groups are arranged in growing order. This arrangement makes the search of words easier.

The Hungarian words are put in a separate place. The sequence of the ordered Russian words determine the order of the Hungarian ones.

One extra cell belongs to every Russian word. This cell contains the different data of the word in question. (For instance, the code of the form class.) One part of this cell is used for storing the address of the adequate Hungarian word.

13 bits are required to determine an address. So 26 bits remain for other data. From this "informational cell" any needed information can be separated by logical multiplication by different constants.

Dictionary look up

First, it shall be determined, how many letters the unknown word consists of. After, the unknown word shall be looked up only in the group of words with given length.

The searching of words is executed by a bisecting method, i.e., if we denote the unknown word by S , and the words of the same length as S , by S_1, S_2, \dots, S_n , and if $S \ll S_n$, so it must be decided, which of the following three formulae holds true:

- $$\begin{aligned} \text{i/ } S &= S_j & /j &= \frac{n}{2}/ \\ \text{ii/ } S &< S_j \\ \text{iii/ } S &> S_j \end{aligned}$$

If ii/ holds true, the same examination must be carried out for $j = j - j/2$, and if iii/ holds true, for $j = j + j/2$.

This procedure must be continued until i/ holds true for any j .

If there does not exist a j which satisfies the above requirement then the next letter from the end of the word is to be cut. The procedure begins all over again.

Morphological analysis

Dictionary look up is at the same time the starting point of morphological analysis. On the basis of the endings which have been cut, the morphological class of the word under consideration can be determined. Some ambiguities can only be solved after a syntactic analysis have been carried out. The checking of programs concerning syntactical analysis has just begun.

BIBLIOGRAPHY OF THE MATERIAL
RECEIVED UP TO JANUARY, 1964

- ANDRON,D., Analyse morphologique du substantif russe, CETA-G, G-101-4 1964.
- AUGEREAU,D., Utilisation des informations sémantiques en traduction automatique, Thèse à l'Univ. de Grenoble.
- AUROUX,A., Contribution à la reconnaissance des structures syntaxiques en traduction mécanique, Thèse à l'Univ. de Grenoble.
- Automatic Language Analysis, Fifth Quarterly Report, Indiana Univ..
- Automatic Language Analysis, Sixth Quaterly Report, Indiana Univ.
- Automatic Language Analysis, Seventh Quarterly Report, Indiana Univ.
- BAR-HILLEL,Y., GAIFMAN,C., SHAMIR,E., On Categorical and Phrase Structure Grammars, Reprint from the Bulletin of the Research Council of Israel.
- BERNARD-GEORGES,A., LAURENT,G., LEVENBACH,D., Analyse morphologique du verbe allemand, CETA-CNLS-CETA, G-503-2
- BLOIS,J., DECRESY,F., MOMMEUS,J., Analyse morphologique automatique du français, Univ. libre de Bruxelles, 1963
- BLOIS,J., MORLET,E., Morphologie du français pur la traduction automatique, Euratom-CCR ISPRA, CETIS 44, 1962
- BUSA,P., R., L'analyse des homographes et des formes flexionnelles, en Cahiers de lexicologie, vol.3, 1961,Paris

- BUSA,R., All non Biblical Dead Sea Scrolls Published up to December 1957 Have Been Indexed, in Sacra Pagina, 1959, Paris-Gembloux
- BUSA,R., Note on Some Lexical and Philisophical Implications of a Computer Symbolic Language, at Symbolic Languages in Data Processing, March 26-31 1962, Rome
- CHOMSKY,N., Formal Properties of Grammars, in Handbook of Mathematical Psychology, vol.II, 1963
- CHOMSKY,N., MILLER,A.G., Finitary Models of Language Users, in Handbook of Psychology, vol.II, 1963
- CHOMSKY,N., MILLER,A.G., Introduction to the Formal Analysis of Natural Languages, in Handbook of Psychology, vol. II., 1963
- Computer-Aided Machine Translation Russian → English, Wayne State Univ.
- DELCROIX,M., EBSTEIN,J., Analyse morphologique du substantif français, CETA-G G-1501-1, 1963
- Dicautom, Univ. libre de Bruxelles, 1963
- Distribution Class, Japan 1963
- DOMONKOS,E., Méthode matematice in realizarea limbii intermediare pentru traducerea automata, in Probleme de Automatizare, IV., Bucuresti
- Étude morphologique des classes de mots invariables et des pronoms de la langue russe, CETA-G, G-104-1, 1962
- GAGNY,A., Analyse morphologique du verbe russe, CETA-G G-103-2, 1962
- GROSS,M., Linguistique mathématique et langage de programmation, Laboratoire Central de l'Armement

- GROSS, M., On the equivalence of models of language used in the field of mechanical translation and information retrieval, Research Laboratory of Electronics MIT
- HAYS, D.G., Studies in Machine Translation: Russian Sentence Structure Determination, The Rand Corporation RM-2538
- HAYS, D.G., Grouping and Dependency Theories, The Rand Corporation P-1910
- HAYS, D.G., Basic Principles and Technical Variations in Sentence-structure Determination, reprint from Information Theory
- Hazard and Race Phenomena in Switching Circuits, Academy of the Rumanian, P.R., Univ. of Bucharest
- HIRSCHBERG, L., Punctuation and Automatic Syntactic Analyses
- HIRSCHBERG, L., Les ponctuation, Univ. libre de Bruxelles
- HIRSCHBERG, L., Ponctuation et analyse syntaxique automatique, Univ. libre de Bruxelles, 1962
- HIRSCHBERG, L., Description formelle des données observables du langage écrit, Univ. libre de Bruxelles
- HIRSCHBERG, L., LINCH, I., Discussions sur l'hypothese de projective, Univ. libre de Bruxelles
- JOHNSON, C.D., Russian Affixal Lexes and Lex Sequences, Univ. of Calif. 1963
- JOHNSON, C.D., The Punching Code for Chinese Text, Univ. of California, 1963
- LAMB, S.M., Current Research of the Machine Translation Project, Univ. of California, Berkeley
- LAMB, S.M., Sememic Approach to Structural Semantics, Univ. of California

- LAMB, S.M., Nature of the Machine Translation Problem, Univ. of California, 1963
- LAURENT, G., BERNARD-GEORGES, A., LEVENBACH, Etude morphologique de l'adjectif allemand, CEEA, ADR-G
- LAURENT, G., BERNARD-GEORGES, A., LEVENBACH, D., Etude morphologique de substantif allemand, CEEA, ADR-G
- OPLER, A., BAIARD, N., Experience in Developing Information Retrieval Systems on Large Electronic Computers, from Proceedings of the International Conference on Scientific Information, 1958
- Proceeding of the Conference of Federally Sponsored Machine Translation Groups on MT-oriented Syntactical Analysis, Summary of Princeton, NJ, June 13-15 1962
- Research in Machine Translation, Russian → English, vol. II, Wayne State Univ., 1962
- RHODES, I., A New Approach to the Mechanical Syntactic Analysis of Russian, from Mechanical Translation
- RHODES, I., ALT, F.L., Hindsight Techniques in MT of Natural Languages Journal of Research of the NBS
- ROUSSE, F., Locutions françaises issues de mots russes uniques, Univ. libre de Bruxelles
- SAKAI, ITIROO., Syntactic Functions and Distribution Classes Defined on the Basis of the Set Theory, Defence Agency, Japan, 1963
- SAKAI, ITIROO., Syntax in Universal Translation, Defence Agency, Japan
- SCHNELLE, H., Programmieren linguistischer Automaten, Institut für Phonetik und Kommunikationsforschung, Bonn
- SESTIER, A., DUPUIS, L., La place de la syntaxe dans la traduction automatique des langues

Slovo a Slovesnost, 1962:1, 2, 1963: 1,2, Praha

TORRE,L., Analyse morphologique de l'adjectif russe, CETA-G,
G-102-3 1962

VAUQUOIS,B., Language artificiels-systemes formels et traduction automatique, CETA-G

VAUQUOIS,B., VEYRUNES,J., Présentation de l'analyse morphologique du russe, CETA-G, C-100-C, 1962

VEILLON,G., Consultation d'un dictionnaire et analyse morphologique en traduction automatique, Thèse à l'Univ. de Grenoble

VEILLON,G., Note sur la codification de l'alphabet allemand, CEEA, ADR-G

VEILLON,G., Présentation de l'analyse morphologique et du programme de dictionnaire allemand, CETA-G, G-500-A, 1963

YAMADE,S., Étude morphologique du Japonais, CET-G, G-1000-2, 1962

АНДРЕЕВ,Н.Д.(ред.), Материалы по математической лингвистике и машинному переводу II., Ленинград, 1963

Машинный перевод и прикладная лингвистика 2-5

Научно-техническая информация, 1963:10, Москва

РАВИЧ,Р.Д., Библиография зарубежных работ по машинному переводу, Москва, 1962

